230/27.8.7

THE BRITISH JOURNAL OF PSYCHOLOGY

EDITED BY
JAMES DREVER

WITH THE ASSISTANCE OF
C. BURT

AIDED IN THE SELECTION OF PAPERS BY

F. C. BARTLETT
E. A. BOTT
R. S. CREED
G. C. DREW
L. S. HEARNSHAW
M. G. KENDALL
T. H. PEAR
G. RYLE
W. J. H. SPROTT
R. H. THOULESS

ASSISTANT EDITOR: IAN M. L. HUNTER

VOLUME XLVIII, 1957

BUREAU OF EDUC. & PSYL. RESEARCH.
LIBRARY.

CAMBRIDGE
AT THE UNIVERSITY PRESS
1957

Printed in Great Britain at the University Press, Cambridge
(Brooke Crutchley, University Printer)
and published by the Cambridge University Press
Bentley House, 200 Euston Road, London, N.W. 1
American Branch: 32 East 57th Street, New York 22, N.Y.
Agents for Canada: The Macmillan Company of Canada Limited

Bureau Edn. sy Research
DAVID H ING COLLEGE
Dated 27.8.70
Accs. No. g 230

CONTENTS OF VOL. XLVIII

PART 1. FEBRUARY 1957

PAGE

Determinants of emotionality in the rat. I. Situational factors. By P. L. Broadhurst 1

The assessment of anxiety as an intervening variable in the delinquent behaviour of M.D. subjects. Galvanic skin response and leg-persistence indices. By John E. Tong 13

Neuroticism in children. I. An investigation of normal and neurotic group differences. By J. G. Thorpe and D. P. James 26

Cognitive inference in perceptual activity. By M. D. Vernon 35

Personality and perception: the constancy effect and introversion. By J. Amor Ardis and Elizabeth Fraser 48

Visual discrimination of orientation by octopus. By N. S. Sutherland . 55

Corrigendum 71

Publications recently received 72

PART 2. MAY 1957

Individual differences in reacting to delayed auditory feedback. By J. T. Beaumont and B. M. Foss 85

Discrimination and operant extinction. By H. M. B. Hurwitz and J. Cutts 90

The psychological refractory phase. By Thomas Marill 93

The normal distribution of intelligence: a critique. By D. G. Lewis . . 98

A one-way window. By C. B. Hindley and Frank Falkner . . . 105

Conflict and choice time. By D. E. Berlyne 106

Personality factors and the rate of conditioning. By Cyril M. Franks . 119

An experimental investigation of sewing-machine skill. By W. T. Singleton 127

A note on the role of 'guidance' in learning. By J. M. von Wright . . 133

Attention to change, conditioned inhibition (${}_{S}I_{R}$) and stimulus satiation. By D. E. Berlyne 138

A note on the fallacy of the 'phenomenological fallacy'. By J. R. Smythies 141

Publications recently received 145

Other publications received 159

PART 3. AUGUST 1957

PAGE

The distribution of intelligence. By CYRIL BURT 161

Corrigendum 175

On the validity of applying communication theory to experimental psychology. By E. COLIN CHERRY 176

On some psychophysical techniques employing a single manipulandum. By LAWRENCE WEISKRANTZ 189

Perceptual and personality dynamics in reactions to ambiguity. By VERNON HAMILTON 200

Frequency, recency, association value and tachistoscopic identification. By G. W. H. LEYTHAM 216

The prediction of examination marks of mechanical engineering students at King's College, Newcastle. By R. F. GARSIDE 219

Shape recognition in rats. By P. C. DODWELL 221

Publications recently received 230

PART 4. NOVEMBER 1957

Individual differences in intentional and incidental learning. By MAVIS PLENDERLEITH and LEO POSTMAN 241

Rubella deaf children: a Rorschach study employing Munroe inspection technique. By D. MARJORIE BINDON 249

Previous knowledge and memory. By E. C. POULTON 259

Measures of subjective probability. By JOHN COHEN, E. J. DEARNALEY and C. E. M. HANSEL 271

The qualitative investigation of intelligence. By ANN SHEDDEN MCCOLSKEY 276

The solving of three-term series problems. By IAN M. L. HUNTER . . 286

The influence of extraneous stimulation on the reaction time of schizophrenics. By J. TIZARD and P. H. VENABLES 299

The boundaries between the person and the world: a note on the paper by Gardner Murphy. By LEON LITWINSKI 306

Publications recently received 307

INDEX OF AUTHORS

PAGE
Ardis, J. A. 48
Beaumont, J. T. 85
Berlyne, D. E. 106, 138
Bindon, D. M. 249
Broadhurst, P. L. 1
Burt, C. 161
Cherry, E. C. 176
Cohen, J. 271
Cutts, J. 90
Dearnaley, E. J. 271
Dodwell, P. C. 221
Falkner, F. 105
Foss, B. M. 85
Franks, C. M. 119
Fraser, E. 48
Garside, R. F. 219
Hamilton, V. 200
Hansel, C. E. M. 271
Hundley, C. B. 105
Hunter, I. M. L. 286
PAGE
Hurwitz, H. M. B. 90
James, D. P. 26
Lewis, D. G. 98
Leytham, G. W. H. 216
Litwinski, L. 306
Marill, T. 93
McColskey, A. S. 276
Plenderleith, M. 241
Postman, L. 241
Poulton, E. C. 259
Singleton, W. T. 127
Smythies, J. R. 141
Sutherland, N. S. 55
Thorpe, J. G. 26
Tizard, J. 299
Tong, J. E. 13
Venables, P. H. 299
Vernon, M. D. 35
Weiskrantz, L. 189
Wright, J. M. von. 133

INDEX OF PUBLICATIONS REVIEWED

Acta Physiologica Scandinavica, Vol. 39, Suppl. 134 235
Anstey, E. & Mercer, E. O. *Interviewing for the Selection of Staff* 315
Bellows, R. M. *Psychology of Personnel in Business and Industry* 316
Bhatia, C. M. *Performance Tests of Intelligence under Indian Conditions* 80
Brill, N. Q. & Beebe, G. W. *A Follow-up Study of War Neuroses* 79
Bruner, J. S., Goodnow, J. J. & Austin, G. A. *A Study of Thinking* 145
Brunswik, E. *Perception and the Representative Design of Psychological Experiments* 146
Bugelski, B. R. *The Psychology of Learning* 237
Burn, M. *Mr Lyward's Answer* 152
Buytendijk, F. J. J. *Theorie der menschlichen Haltung und Bewegung* 80
Caldwell, E. M. *A Case of Spatial Inability in a Cerebral Palsied Child* 234
Caligor, L. *A New Approach to Figure Drawing* 319
Carroll, J. B. (ed.). *Language, Thought and Reality* 319
Chauchard, P. *Les Mécanismes Cérébraux de la Prise de Conscience* 82
Cherry, C. (ed.). *Information Theory* 313
Cherry, C. *On Human Communication* 315
Cholden, L. (ed.). *Lysergic Acid Diethylamide and Mescaline in Experimental Psychiatry* 239
Clark, M. M. *Left-handedness* 318
Cronbach, L. J. & Gleser, G. C. *Psychological Tests and Personnel Decisions* 312
David, H. P. & Bracken, H. von (eds.). *Perspectives in Personality Theory* 310
Eng, H. *The Psychology of Child and Youth Drawing* 233
Eysenck, H. J., Granger, G. W. & Brengelmann, J. C. *Perceptual Processes and Mental Illness* 318
Festinger, L., Riecken, H. W. & Schachter, S. *When Prophesy Fails* 157
Flugel, J. C. *Studies in Feeling and Desire* 150
Fraisse, P. *Les Structures Rythmiques* 236
Getz, S. *Environment and the Deaf Child* 151
Ghiselli, E. H. & Brown, C. W. *Personnel and Industrial Psychology* 154
Gilman, G. *Human Relations in the Industrial Southeast* 78
Glover, E. *On the Early Development of Mind* 320
Godwin, G. *Crime and Social Action* 84
Guilford, J. P. *Fundamental Statistics in Psychology and Education* 75
Hall, C. S. & Lindzey, G. *Theories of Personality* 309
Halmos, P. *Towards a Measure of Man* 238
Hearing Defects in School Children 151
L'Instinct dans le Comportement des Animaux et de l'Homme 239
Israel, J. *Self-evaluation and Rejection in Groups* 80
Jersild, A. T. *The Psychology of Adolescence* 233
Kendall, M. G. *Rank Correlation Methods* 77
Klein, J. *The Study of Groups* 232
Klopfer, B. *et al. Developments in the Rorschach Technique*, Vol. II 150
Köhler, W. *The Mentality of Apes* 237
Krout, M. H. (ed.). *Psychology, Psychiatry and the Public Interest* 158

Lafitte, P. *The Person in Psychology* 238
Leary, T. *Interpersonal Diagnosis of Personality* 230
Lidell, H. S. *Emotional Hazards in Animals and Man* 82
Maier, N. R. F., Solem, A. R. & Maier, A. A. *Supervisory and Executive Development* 316
Mainwaring, J. *Psychology in the Classroom* 233
Malzberg, B. & Lee, E. S. *Migration and Mental Disease* 84
Mannheim, H. & Wilkins, L. T. *Prediction Methods in Relation to Borstal Training* 231
Michaelis, A. R. *Research Films in Biology, etc.* 236
Miner, J. B. *Intelligence in the United States* 317
Morgan, C. T. *Introduction to Psychology* 77
O'Connor, N. & Tizard, J. *The Social Problem of Mental Deficiency* 148
Oeser, O. A. & Hammond, S. B. (eds.). *Social Structure and Personality in a City* 147
Roe, A. *The Psychology of Occupations* 154
Rose, A. M. (ed.). *Mental Health and Mental Disorder* 156
Ruby, L. *The Art of Making Sense* 82
Scheler, M. *The Nature of Sympathy* 83
Schultz, J. H. *Das autogene Training* 158
Sechehaye, M. *A New Psychotherapy in Schizophrenia* 81
Shaffer, L. F. & Shoben, E. J. *Psychology of Adjustment* 155
Sholl, D. A. *The Organization of the Cerebral Cortex* 72
Siegel, S. *Non-parametric Statistics for the Behavioural Sciences* 152
Simon, B. (ed.). *Psychology in the Soviet Union* 311
Smith, M. B., Bruner, J. S. & White, R. W. *Opinions and Personality* 78
Smythies, J. R. *Analysis of Perception* 74
Spence, K. W. *Behaviour Theory and Conditioning* 230
Stagner, R. *The Psychology of Industrial Conflict* 78
Stott, D. H. *Unsettled Children and their Families* 153
Strachey, A. *The Unconscious Motives of War* 237
Strauss, A. (ed.). *The Social Psychology of G. H. Mead* 155
Tanner, J. M. & Inhelder, B. (eds.). *Discussions on Child Development*, Vol. I 76; Vol. II 232
Thorpe, W. H. *Learning and Instinct in Animals* 73
Tredgold, R. F. & Soddy, K. *A Text-book of Mental Deficiency* 149
Vernon, P. E. *The Measurement of Abilities* 75
Vernon, P. E. (ed.). *Secondary School Selection* 307
Wallen, R. W. *Clinical Psychology* 234
Ward, R. H. *A Drug-taker's Notes* 239
Way, L. *Alfred Adler* 76
Wenger, M. A., Jones, F. N. & Jones, M. H. *Physiological Psychology* 235

VOLUME 48 FEBRUARY 1957 PART 1

DETERMINANTS OF EMOTIONALITY IN THE RAT

I. SITUATIONAL FACTORS

By P. L. BROADHURST

University of London, Institute of Psychiatry (Maudsley Hospital)

An experiment was designed to establish the optimum conditions for using Hall's open-field test of emotionality in the white rat. A large ($N = 192$) factorial design permitted simultaneous evaluation of the following components of the test which were thought to be productive of emotional responses: size of test arena, intensity of illumination and intensity of sound. The effects of pre-trial shock, food deprivation and sex difference were also investigated. Other variables were strictly controlled.

The results, principally analysed in terms of defecation and ambulation scores, show that females defecate significantly less than males and ambulate more, that noise is a potent factor in evoking emotional elimination and light less so. Arena size does not affect defecation, but rats run farther in a larger arena. Pre-trial shock reduces defecation, whereas food deprivation has no significant effect. Interactions are few, and mostly concern sex differences in response.

It is concluded that the test can yield sensitive indicators of an emotional response, and that this response is susceptible, within limits, to experimental control. The implications for the use of the test are discussed.

I. INTRODUCTION

Loss of control over elimination may be induced by fear. This has been shown (Stouffer, Guttman, Suchman, Lazarsfeld, Star, Shirley & Clausen, 1950, p. 142) to be the extreme fear symptom on the continuum of physiological symptoms reported by soldiers as occurring in battle conditions. Similarly, the rat will defecate or urinate or do both when placed in a strange, fear-provoking situation. As Munn (1950, p. 98) notes in his review of the work on emotional behaviour in the rat, observation of this tendency has long been used as a measure of emotionality (Yerkes, 1913; Stone, 1932). Systematic work began with the validation studies of Hall (1934, 1936*a*, *b*), who used a circumscribed arena or open field in which to observe it. His demonstration of the validity of defecation as an emotional index showed that it decreases with successive exposures to the same situation and has been repeatedly confirmed (e.g. Biel & O'Kelly, 1940; Farris & Yeakel, 1945). Factorial validity has also been demonstrated (Billingslea, 1942). A negative correlation between elimination scores and ambulatory activity in the open field is a feature of the results derived from this test (Hall, 1936*b*).

But the technique has been criticized because it is space-consuming (Lebo, 1953), because of the failure to obtain interpretable amounts of defecation (Griffiths & Stringer, 1952), and because such scores, when obtained, do not always correlate with other single purported measures of emotionality or timidity (O'Kelly, 1940; Bindra & Thompson, 1953; Hunt & Otis, 1953). The last authors suggest that the test should be more stressful. It is clear, therefore, that the fear-provoking components of the test should be identified and standardized, preferably while retaining the simplicity of Hall's original procedure. Accordingly, the present experiment has been designed to allow us to examine, independently, each of several components of the open-field test thought to be stressful, and to evaluate the effects of specific changes in their intensity.

Furthermore, in much animal experimentation, what might be called 'background conditions, e.g. the strain of the subjects, their age, previous experience and husbandry, are not specified (Christie, 1951). Recent evidence shows that such conditions may well be determinants of behaviour in this test (Watson, 1954; Weininger, 1954). Thus, while the purpose of this study is to attempt the definition of the optimum conditions for using the test, the effects of background conditions cannot be ignored. It is proposed, therefore, to maintain such conditions constant for all animals, and to concentrate attention on the variables inherent in the testing situation itself, except for those which are imposed (sex and litter membership), or which it seems profitable to vary (hunger drive).

Illumination of the arena is the only test condition which has been systematically varied, Evans & Hunt (1942) finding that defecation scores increased with brightness. Noise has been employed as an emotional stimulus (Patrick, 1931; Bindra & Thompson, 1953); and its use is advantageous since the same noise at a lower intensity can serve as a sound screen, thus avoiding the need for sound-proofing the subjects against distracting external noises. The size of the test arena used has varied greatly, but, on *a priori* grounds, one might expect a larger arena to be more stressful for the somewhat agoraphobic rat. Methods of making the open-field test even more stressful for the rat (Anderson, 1939, 1940) are difficult to adapt to a large arena. The procedure adopted here is that of the 'pre-trial shock', a technique which may be expected to heighten anxiety. In addition to these four variables within the testing situation, it is proposed to investigate the effect of hunger drive, because Hall (1936*a*) found hungry rats to be less emotional. His claim that this effect could not be due to a smaller available bulk of faecal material (since urination scores were also affected) is invalidated by the evidence that food-deprived rats also drink less (e.g. Verplanck & Hayes, 1953). The evidence relating to sex differences in open-field test scores is both scanty and contradictory (Hall, 1934; Parker, 1939; Farris & Yeakel, 1945; Thompson & Bindra, 1952).

Certain predictions regarding the effect on open-field test scores of the experimental variables can be made, bearing in mind that increased elimination and decreased activity may be regarded as characteristic of the more emotional rat (Hall, 1936*b*). They may be summarized as follows:

$$\left.\begin{array}{l}\text{larger arena}\\ \text{brighter light}\\ \text{louder noise}\\ \text{pre-trial shock}\end{array}\right\}\ \text{will result in greater elimination and less activity,}$$

food deprivation will result in less elimination and more activity.

It will be appreciated that these expectations are tentative and have heuristic value only. In view of conflicting evidence regarding sex difference in scores, no predictions are made for the effects of this variable.

II. Experiment

(i) *Experimental design*

The selection of an analysis of variance design is determined by several considerations, prominent among which are the number of independent variables to be manipulated and the number of subjects available. In the present experiment, it is desired to evaluate the effect of five variables on the subjects' behaviour in the open-field test. To these five must be added two more, sex and litter membership, which are imposed. About 200 animals were available for the experiment. Since such a large number cannot be given the test on any one occasion, the variables of the subjects' age and the time of day of testing are inevitably introduced.

Five variables, each at two levels, plus sex, requires sixty-four animals, in order to assign one to each unique combination of treatments. The primary choice lay between using the remaining subjects for increasing the number of levels of treatments like intensity of light, noise, etc., and foregoing the advantages of replicating the design, or assigning more than one subject to combinations of treatment at two levels only. The latter alternative was chosen. Replication permits evaluation of any effects of age obtainable by varying systematically the mean age of the subjects assigned to each replication, within, that is, the small age range of 100–110 days here employed. Any effects of the time of day of testing—morning sessions as opposed to afternoon—can also be determined.

In order to investigate the effect of inter-litter differences, a design involving blocks was required. The one selected is the cubic lattice of Yates (1939). It is so called because each of the three replications requires k^3 subjects, where k equals the number of subjects in each block. In this experiment, blocks represent litters and $k=4$. In only seven cases (out of forty-eight) was one non-litter-mate required to complete a block. On the other hand, this equating of litters with blocks was necessarily imperfect in that we had only twenty-four litters to be distributed over the forty-eight blocks, and so some litters were distributed over two or even three blocks. The sixty-four combinations of treatments were randomly assigned to block positions which are balanced between each replication. The subjects were allocated to replications by age, to blocks by virtue of litter composition, and to cages (and hence to cage mates) at random within sex. Where this resulted in fewer than three rats in a cage, one or two non-experimental animals of the same age and like sex were used to make up the standard cage complement of three. The order of testing was determined by block membership, and the first eight blocks in each replication (thirty-two animals) were given the open-field test in the morning, the next eight in the afternoon.

(ii) *Apparatus*

Two sizes of open-field arenas were used in order to provide the two levels of treatment required for this variable. The larger arena was made of curved metal units which enclosed, when assembled, an area 6 ft. in diameter with a wall 12½ in. high, the interior surface of which was painted flat white. It was bottomless and stood directly on a granolithic floor on the outermost of six concentric circles divided into segments by radial lines. These floor markings were needed to provide compartments of roughly equal size so that the ambulation could be easily scored as an animal moved around in the arenas.

The smaller arena was designed to go, when in use, inside the large arena and to span the three innermost circles. It had a diameter of 32¾ in. and was made of plywood for ease of handling, but was otherwise identical with the large arena. Around the sides and over the top of the large arena, a light metal structure supported a screen of muslin which served both as a one-way screen and to diffuse the illumination from above. The structure also supported the batteries of lamps and loudspeakers for providing the light and noise stimulation.

The intention was to have this stimulation distributed as uniformly as possible over the arenas. The light was provided by six photographic lamps wired in parallel and arranged at a height of 44 in. in a pattern that produced the most uniform illumination at floor level in either arena. Measurements with a Ferranti light meter (type LMB) showed a range of 170–195 foot-candles in the large arena, and 145–185 f.c. in the small arena. This constituted the high level of illumination. Switching a large resistor in series with the lamps reduced the intensity of illumination to 30–37 f.c. in the large arena and to 22–32 f.c. in the small. This constituted the low level of illumination. While these determinations were being made and during subsequent testing, all other illumination was excluded.

Continuous 'white noise', that is, noise of all frequencies at once, was fed in, via an amplifier, to six loudspeakers, 6½ in. diameter mounted 51 in. above the floor. The source for the white noise was the output of a small commercial neon lamp (Pascal & Swenson, 1952). Appropriate resistors in the circuit (a potential divider) gave two levels of noise. The intensity of these two sound fields was measured by exploring the floor level of the arenas with a calibrated microphone (Acos, type MIC 2), the output of which was amplified and displayed on an oscilloscope (Cossor, model 1049). The wave form was observed to be characteristic of that for white noise, and the sound intensities, as measured in both arenas, were 94 db. (ref. 0·0002 dyne/sq.cm.) for the high level of sound and 87–88 db. for the low level. This softer noise served as the sound screen for masking incidental external noises. The loud noise was below the intensity which Morgan & Galambos (1942) found would cause audiogenic seizures with low frequencies. As will be seen later, there were no significant interactions, in any of the scores, between the noise and light variables and arena size. This suggests that the lack of complete uniformity in the light and sound fields had no profound influence on scores.

The carrying box, in which half the animals were shocked immediately before being tested in the open field, followed the design of Warner (1932). Each wall and half the floor constituted one electrode, and was made from one piece of aluminium, bent to shape. The two pieces were close enough together to force an animal to stand partly on each, yet with sufficient gap between the floor sections to prevent shorting caused by excrement. A Perspex lid was rigidly fixed at right angles to the top of one wall which swung outwards about pivots near the base of the box. This arrangement afforded a wide opening by which to insert and remove the animal, since both lid and one wall moved aside together. Through the feet of a rat standing in the box a current of predetermined value could be switched. This current was obtained from a 2 kV. transformer; it was limited by a 2-megohm series resistor and determined by a variable resistor in series with the transformer primary. Variations in the resistance of individual animals, which are known to be large, thus had little influence on the intensity of the shock. Indeed, decreasing the calibrating resistance of 220 K. to zero by shorting the electrode terminals altered the current by less than 11%.

(iii) *Subjects*

A group of albino rats were bred, as subjects for the experiment, from a remote Wistar strain, obtained originally from a dealer. It had been maintained by random mating in this laboratory for five generations. Twenty-four litters comprising 96 males and 110 females—an effective total of 192 with equal numbers of each sex—were reared under strictly standardized conditions from birth to the time of testing. An automatically controlled lighting system supplemented daylight in such a way that the colony room was in darkness only during the period 9.0 p.m. to 3.30 a.m. G.M.T. Handling was kept to the minimum. The litters were weaned at 21 days after birth, when the pups were sexed and caged in groups of five. At 50 days their ears were clipped under ether anaesthesia for further identification and they were caged thereafter in threes. The cages measure 9 by 9 by 8½ in. high and have solid metal backs, tops and sides. An *ad libitum* diet of rat cubes (MRC diet no. 41) and tap water was provided. All subjects were between 100 and 110 days old when experimentation on them began; at this time they had been handled only five times since birth. Thus, not only were the subjects experimentally naïve at the beginning of testing, but also they had not undergone any taming or 'gentling'.

(iv) *Procedure*

Food deprivation was commenced 5 days after the final cage allocations were made and 13 days before testing began. It was established by degrees, only reaching the final period of 2 hr. feeding per day after 9 days. This lengthy procedure was adopted in order to minimize weight loss and to ensure stability of hunger drive before the experiment began (Hall, Smith, Schnitzer & Hanford, 1953).

The 2 hr. feeding time followed the morning and the afternoon sessions, so that the food-deprived subjects were tested after a period ranging from 19½ to 21½ hr. without food. The other half of the subjects, who were not deprived of food, were subjected to disturbance comparable to that occasioned by replacing, and later removing, the food-hoppers from the cages of the animals actually deprived.

Animals were individually transported in the carrying box, described above, from their cages in the colony room to the open-field test where they were detained in the box for 15 sec. The pre-shock group were subjected to a current of 0·25 mA. a.c. through the feet during this time. This shock was always sufficient to evoke running movements and squeaking almost always. Defecation was less frequent. The combination of noise and light levels and arena size appropriate to the particular subject having been established, the rat was placed in the arena half-way between edge and centre, facing outwards. Each subject was left in the arena for 2 min. during which time ambulation was noted on a prepared plan of the arena floor, together with the incidence and amount of defecation. A urination score was not used because of the difficulty of scoring this variable objectively. Even without any attempt at quantification, and merely scoring urination as present or absent, it is almost impossible, with present techniques and especially in the large arena, to be sure that one spot of urine deposited somewhere has not been overlooked. The contrast with the ease of scoring defecation by counting the number of faeces deposited is noteworthy. Testing took place on five successive days. The order in which individuals were tested was not varied. The three replications of the experiment were begun at intervals of 1 week.

III. Results

(i) *Scoring*

The defecation scores were analysed in two ways: first, by the frequency of defecation, that is, the number of days out of the five upon which any defecation occurred; and secondly, in terms of the amount of defecation, that is, the number of boluses deposited when defecation did occur. This was done in order to obtain a score which reflected the amount of defecation independent of its frequency. Ambulation, obtained as a raw score of the number of subdivisions marked on the floor which the subject entered, was converted into metres by multiplying by appropriate constants.

In each case a total or an average score for the 5 days was used in the analysis, because examination of the day-to-day totals showed no evidence of differences in trend attributable to treatment effects. But it is of interest to note that the data from all 192 subjects (irrespective of treatment), for the average daily defecation score, clearly shows the characteristic decline in magnitude with repeated daily exposure of the subjects to the open-field test. In contrast, the average daily ambulation score shows little change after an initial drop. It is clear that there is no pronounced day-by-day trend towards the increased ambulation to be expected if familiarity with the situation decreased a tendency to emotional 'freezing'.

(ii) *Litter groupings, age and time of testing*

Before proceeding further it is desirable to indicate how far it is possible to eliminate from consideration possible sources of bias in the results due to litter grouping, age and time of day of testing. A preliminary investigation of litter groupings was made by considering the results in the form of block totals. An analysis of variance for each of the three scores analysed (that is, defecation frequency, defecation amount and ambulation) showed a significant 'between-blocks' effect, as might be expected as a result of the differences in the experimental treatment imposed on their members. Yet, only those blocks in which most of the animals had been subjected to treatment combinations which included significant main effects (see later) reflected this circumstance by the extremeness of their scores. Moreover, there was no evidence, among those blocks representing parts of the same litter, of any systematic high or low scoring. It was concluded from this simple analysis that litter differences could be excluded from further consideration, and that the more elaborate treatment which this design allows (Yates, 1939) was not needed.

It will be recalled that age was confounded with replication in that there was a small, though highly significant, difference between the mean ages of the subjects at the start of the three replications—102, 105 and 108 days respectively. Yet, for none of the scores analysed was there a significant F-ratio for the replication variable. This shows that such small differences in age do not affect the responses to the open-field test, and permits us to pool the results of the three replications. In subsequent discussion of the results this has been done.

The mean score for the morning session was compared with that for the afternoon session for each of the three scores. In each case the difference was small and non-significant. Thus, time of day of testing did not affect the results.

(iii) *Main findings*

Table 1 shows the mean value for each of the two sub-groups ($n=96$) comprised in each of the six treatments. The results of the analyses of variance* performed on the data are also presented in summary form. Opposite each treatment the P value, showing the significance of the associated F-ratio, is given. The F-ratios were obtained by testing the mean sums of squares either against the remainder error variance estimate, with which the non-significant interactions had been pooled, or against the largest of the significant interactions found. Reference to the column giving the degrees of freedom for each F-ratio will show which kind of error term was used in any particular case. It will be noted that, in a design having two levels for each treatment, both main effects and interactions, of no matter what order, will have only one degree of freedom. The simple interactions which turned out to be significant in the analysis of defecation frequency and amount scores are shown in Tables 2 and 3. The results of the t-tests of the significance of the

Table 1. *Mean scores for sub-groups and summary of results of analyses of variance*

		Defecation frequency			Defecation amount			Ambulation		
Treatment	Sub-group	Means (days)	Signif. of F-ratio	D.F.	Means (boluses)	Signif. of F-ratio	D.F.	Means (metres)	Signif. of F-ratio	D.F.
Sex	Males	3·8	$P<0{\cdot}001$	1, 182	3·2	N.S.	1, 1	8·4	N.S.	1, 1
	Females	2·0			1·8			11·4		
Noise	Loud	3·4	$P<0{\cdot}001$	1, 182	2·8	N.S.	1, 1	*11·2*	$P<0{\cdot}001$	1, 173
	Soft	2·4			2·2			*8·6*		
Light	Bright	3·2	N.S.	1, 1	2·7	$P<0{\cdot}05$	1, 182	*10·4*	N.S.	1, 1
	Dim	2·7			2·3			*9·4*		
Arena size	Large	*2·7*	N.S.	1, 182	*2·4*	N.S.	1, 182	*12·5*	N.S.	1, 1
	Small	*3·1*			*2·5*			*7·3*		
Food deprivation	Not hungry	3·1	N.S.	1, 182	2·6	N.S.	1, 182	9·9	N.S.	1, 173
	Hungry	2·7			2·4			9·9		
Pre-trial shock	Shocked	3·0	N.S.	1, 1	*2·2*	$P<0{\cdot}01$	1, 182	9·3	N.S.	1, 1
	Not shocked	2·9			*2·7*			10·5		
Overall means		2·9			2·5			9·9		
Error variance estimate		2·4			1·2			14·6		

Figures *in italics* indicate differences in the non-predicted direction (see text).

Table 2. *Defecation frequency score: interaction between the light and pre-trial shock treatments*

		Light		
		Bright		Dim
	Shocked	3·5	$P<0{\cdot}001$	*2·4*
Pre-trial shock		$P<0{\cdot}01$		N.S.
	Not shocked	*2·8*	N.S.	*3·0*

Figures *in italics* indicate pairs of mean scores differing in the non-predicted direction (see text).

---

* The usual tests of the assumptions of homogeneity of variance and normality of distribution underlying the analysis of variance are not readily applicable to this design. Examination of the data did not reveal striking departures from these assumptions; moreover, there is now evidence that the F-test is insensitive to moderate departures from them (Lindquist, 1953, p. 78f.)

differences between the various pairs of cell means are indicated in the appropriate positions between the pairs of scores. All levels of significance were obtained from a one-tail distribution, except in the case of sex difference (where no prediction had been made), and in those cases where a difference in the non-predicted direction was found. Italics are used in the tables to indicate these latter. Such differences are, of course, not significant in the predicted direction; accordingly, a two-tail distribution was used to evaluate the observed differences.*

Table 3. *Defecation amount scores: interaction between the noise and sex variables*

		Noise		
		Loud		Soft
Sex	Males	3·3	N.S.	3·1
		$P<0{\cdot}001$		$P<0{\cdot}001$
	Females	2·3	$P<0{\cdot}001$	1·2

The analysis of variance of the ambulation scores is complicated by a significant third-order interaction which cannot be conveniently tabulated. Study of this interaction shows that the light and pre-trial shock factors vary apparently unsystematically with respect to each other and to the other two factors involved, arena size and sex. Arena size has a consistent influence on ambulation scores in the non-predicted direction, that is, rats always ambulate more in the larger arena. This permits the application of a *t*-test to the difference between the means which yields the highly significant value of $P < 0{\cdot}001$. The sex variable shows a similar effect; females ambulate very significantly more than males ($P < 0{\cdot}001$). The simple interaction between these two variables is given in Table 4.

Table 4. *Ambulation scores: interaction between the arena size and sex variables*

		Arena size		
		Large		Small
Sex	Males	10·3	*P<0·001*	6·4
		$P<0{\cdot}001$		$P<0{\cdot}05$
	Females	14·6	*P<0·001*	8·2

Probabilities *in italics* indicate pairs of mean scores differing in the non-predicted direction (see text).

Several points arise from a consideration of these results. The sex difference which emerges from our data is unequivocal. Males defecate significantly more in both frequency and amount and ambulate significantly less than females. This finding illustrates the frequently reported negative correlation between defecation and activity scores in the open-field test; the emotional animal defecates more and ambulates less than the non-emotional. Another aspect of the sex difference is seen in the interaction Tables 3 and 4, each of which shows that females are more responsive to one of the situational variables. Noise is the only other factor which produces an effect on scores comparable to sex in size and consistency. This potent effect obscures their usual relationship, for, while noise in general evokes greater defecation, it also increases rather than decreases ambulation. Intensity of illumination is a less powerful factor in the production of emotional responses. The higher level has no consistent effect on ambulation, but does yield higher defecation amount scores, though at a low level of significance. It has a marked effect upon defecation frequency scores, but only if the animals have been previously shocked (Table 2). This suggests that the shock sensitizes subjects to the effects of light. Greater arena size is ineffective in provoking more defecation. The effect is confined to the ambulation scores where it is very marked. The effect of food deprivation is a very small one. This absence of significant effect shows defecation scores to be independent of differences in the amount of faecal material available which may have resulted from the food deprivation variable. Similarly, ambulation scores are independent of hunger drive.

---

* The writer is aware that this is a controversial procedure. Justification is claimed in this case because of the tentative nature of the predictions made. It may also be added that the use of a two-tail distribution throughout would not have altered materially the levels of significance reported, or the conclusions drawn from them.

On the other hand, both satiated and hungry rats defecate significantly less in amount if shocked first (Table 1). That is, a reduction in the amount of faecal material available, resulting from previous defecation elicited by pre-trial shocking, does, in fact, have an effect on defecation scores in the open-field test. This finding appears to conflict with the statement above regarding the independence of defecation from faecal bulk. But the apparent paradox may be resolved by considering the pre-trial shock as an independent emotionally stressful experience, and considering the defecation responses to it of the satiated and hungry rats separately. No significant difference is found; that is, in the shocking box, as in the open field itself, possible group differences in faecal bulk are unimportant. But this conclusion must be limited to the first stressful situation imposed on the animal. It can thus be seen that the effect of shock upon the open-field test defecation scores has been obscured by an artefact of the experiment in that shock elicited the same response as did the test itself, and that the lower defecation amount scores of the shocked group is a direct consequence of the previous stressful situation, one which the non-shocked group had avoided. The fact that the shock had no such general effect on frequency of defecation in the open field, but only on its amount, appears to confirm this view.

IV. Discussion

To what extent do the striking and consistent sex differences noted represent a real difference in emotionality between the sexes? We can assume that the found difference in the amount of defecation score represents a difference in the weight of faecal boluses produced (Anderson & Anderson, 1938). But since male rats are heavier than females, the alternative explanation that the difference in amount of defecation merely reflects the sex differences in body weight must be considered (Hall, 1934). The present finding that males also show greater frequency of defecation, taken in conjunction with Billingslea's report (1940) of a near-zero correlation between this variable and body weight, suggests a genuine difference in emotionality. Also, the results of the open-field tests (given, in connexion with another experiment, to a group of twenty-four pre-pubescent rats aged 50 days) show no significant difference in defecation scores, while a weight difference, significant beyond the 1 % level, was found. Apparently the sex difference in emotionality develops after the age of 50 days and before 100 days—the age of the subjects used in this experiment—along with the differentiation of sexual function, and independently of the already established sex difference in weight. Furthermore, the direct test of the relation between the defecation amount scores and the body weight of the sixty-four animals in the third replication of the present study yielded correlations between body weight after testing and the amount of defecation of 0·2 and 0·3 for males and females respectively—low and hardly significant. In sum, therefore, it may be said that while it is not possible to reject completely an hypothesis attributing the difference in the amount of defecation found to a sex difference in body size, the available evidence points away from it. The decrease in defecation in the open field by female rats during periods of sexual receptivity (Anderson, 1940) is a possible contributory cause. Since the oestrous cycle in the female rat lasts 4–5 days, during the five consecutive days of testing all of our female subjects should have been on heat at least once.

These highly significant sex differences in emotional elimination are, therefore, a principal finding of this study. They differ from some previous reports (Farris & Yeakel, 1945; Bindra & Thompson, 1953). Turning to ambulatory activity, Thompson's (1953) finding that females are more active than males is confirmed at the same high level of significance. A greater reactivity of females to situational stimuli has also been observed. This was particularly evident in the defecation response to noise, but a similar effect of the larger arena on ambulation scores was found. All these considerations make it clear that work on individual differences in emotionality in rats should employ different criteria for each sex.

Evans & Hunt's results (1942), relating to the effects of light on defecation scores, are, in general, confirmed and extended to include reaction to noise. Within the limits of this experiment, greater intensity of such physical stimuli yields higher defecation scores. This confirmation of the quasi-lawful relationship between stimulation and the most important index of emotionality in the rat is encouraging—and the measure of experimental control over it, thus achieved, promises well for the future use of the test. Moreover, it supports the status of defecation as a valid measure of emotionality in that variations in putative emotional stimuli did, in fact, influence the measure in the expected direction. But similar expectations regarding the variable of pre-trial shock and arena size yielded no such results. The effect of shock was, in fact, in the reverse direction. However, this apparent discrepancy, as has already been pointed out, can be satisfactorily explained. The large arena did, however, fail to influence defecation in the expected direction. In fact, the trend of results was in the opposite sense; the smaller arena eliciting more, rather than less, defecation (see Table 1). This effect may well be bound up with the large influence it exerted upon ambulation scores. Be that as it may, the finding has practical importance—there is no need to use an arena as large as that introduced by Hall (1934). The justification for the current use of smaller arenas for testing rats (e.g. Thompson & Bindra, 1952; Watson, 1954) thus becomes explicit, and Lebo's argument (1953)—for abandoning the open-field test because it is cumbersome—loses point.

The ambulation score is of less value than defecation as an index of emotionality. The absence of a marked trend towards increased ambulation with repeated testing has already been noted. This is consistent with the findings of other workers (e.g. Berlyne, 1950). Both the larger arena and the louder noise caused an increase, rather than a decrease, in ambulation. In the former case, it is clear that we are dealing with an aspect of exploratory drive—a larger arena causing more ambulation. Montgomery's finding (1951) that the amount of exploration done by rats increased with the size of the maze, and is proportional to the total area, is comparable. In the case of noise, the possible influence of an audiogenic seizure mechanism must not be overlooked. The louder noise was of sufficient intensity to provoke much activity, either directed towards escaping from its intensity, or because the high-frequency components of it were above the threshold for eliciting a motor response which could have led (though never did lead) to an audiogenic seizure. A practical implication seems to be that it is probably better to avoid such intense noise in routine testing, particularly in view of the greater convenience of the softer noise. These findings illustrate a difficulty, noted by Anderson (1940), regarding the argument for the validity of a particular response as an index of emotionality on the grounds that emotional stimuli increase that response. Obviously the possibility of any direct reflex action must be excluded, for, should such an effect be in the same direction as the expected emotional response, a spurious conclusion might follow. Fortunately in this investigation, the effect of noise on ambulation was in a direction suggestive of lower rather than higher, emotionality. And with respect to the conclusions reached regarding defecation, it is difficult to envisage any direct effect of either noise or light upon faecal evacuation.

The relative absence of any effects attributable to hunger are of considerable interest. Hall's report (1936*a*) of differences in emotional responses between hungry and satiated rats is not confirmed and, accordingly, his argument that adjustment to the open-field situation is a function of hunger drive is not supported. Regarding the effect of food deprivation upon ambulation scores, the literature contains similar results (Montgomery, 1952; Thompson, 1953). Taken in conjunction with the relation between arena size and ambulation, this finding that increased hunger drive (which is known to augment significantly general activity in cages) has no significant effect on ambulation, compels the conclusion that the activity which the open-field test measures is externally rather than internally stimulated. That is to say, we are dealing here not with general activity, increased by hunger, but with exploratory activity stimulated by novel environmental conditions. This view has already been advocated by Anderson (1938), who notes that the lack of correlation between open-field ambulation and activity, as measured in cages, supports the distinction. Further support may be derived from the conclusions of Montgomery (1953), Berlyne (1950) and Harlow (1953), all of whom describe behaviour apparently motivated by exploratory drive.

One further point is worthy of comment. The relative success, attained in this experiment, of obtaining measurable amounts of defecation confirms the view (Hunt & Otis, 1953) that the open-field test might function more satisfactorily when made more stressful, and contrasts with those studies in which the test, used under conventional conditions, has yielded little or no score. But the strict maintenance of 'experiential naïveté' (Christie, 1951) in the subjects may also have been important. The effect on test scores of varying such background or life-history factors, held constant in the present study, is under investigation.

To summarize the practical implications for the use of the open-field test: the results suggest that a small, brightly lit arena with a noise level at least loud enough to act as a sound screen will yield satisfactory results. It is immaterial whether or not the subjects are hungry, but their sex is important, as is the occurrence in their immediate past of a markedly stressful experience. Defecation is the emotionality index of choice, measured preferably by amount rather than frequency, since the former will yield, if necessary, a continuous distribution of scores from a single testing and is slightly more sensitive.

I am indebted to Prof. H. J. Eysenck, who proposed this experiment, for many helpful suggestions. The assistance of Mr A. S. C. Ehrenberg with the design and statistical analysis, and of Mr W. C. R. Withers with the apparatus is gratefully acknowledged.

REFERENCES

ANDERSON, E. E. (1938). The interrelationship of drives in the male albino rat. II. Intercorrelations between 47 measures of drives and of learning. *Comp. Psychol. Monogr.* **14**, no. 72, pp. 1–119.

ANDERSON, E. E. (1939). The effect of the presence of a second animal upon emotional behavior in the male albino rat. *J. Soc. Psychol.* **10**, 265–8.

ANDERSON, E. E. (1940). The sex hormones and emotional behavior. I. The effects of sexual receptivity upon timidity in the female rat. *J. Genet. Psychol.* **56**, 149–58.

ANDERSON, E. E. & ANDERSON, SARAH F. (1938). The relation between the weight of the endocrine glands and measures of sexual, emotional, and exploratory behavior in the male albino rat. *J. Comp. Psychol.* **26**, 459–74.

BERLYNE, D. E. (1950). Novelty and curiosity as determinants of exploratory behaviour. *Brit. J. Psychol.* **41**, 68–80.

BIEL, W. C. & O'KELLY, L. I. (1940). The effect of cortical lesions on emotional and regressive behavior in the rat. I. Emotional behavior. *J. Comp. Psychol.* **30**, 221–40.

BILLINGSLEA, F. Y. (1940). The relationship between emotionality, activity, curiosity, persistence and weight in the male rat. *J. Comp. Psychol.* **29**, 315–25.

BILLINGSLEA, F. Y. (1942). Intercorrelational analysis of certain behavior salients in the rat. *J. Comp. Psychol.* **34**, 203–11.

BINDRA, D. & THOMPSON, W. R. (1953). An evaluation of defecation and urination as measures of fearfulness. *J. Comp. Physiol. Psychol.* **46**, 43–5.

CHRISTIE, R. (1951). Experimental naïveté and experiential naïveté. *Psychol. Bull.* **48**, 327–39.

EVANS, J. T. & HUNT, J. McV. (1942). The 'emotionality' of rats. *Amer. J. Psychol.* **55**, 528–45.

FARRIS, E. J. & YEAKEL, ELEANOR H. (1945). Emotional behavior of gray Norway and Wistar albino rats. *J. Comp. Psychol.* **38**, 109–18.

GRIFFITHS, W. J., JNR. & STRINGER, W. F. (1952). The effects of intense stimulation experienced during infancy on adult behavior in the rat. *J. Comp. Physiol. Psychol.* **45**, 301–6.

HALL, C. S. (1934). Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. *J. Comp. Psychol.* **18**, 385–403.

HALL, C. S. (1936*a*). Emotional behavior in the rat. II. The relationship between need and emotionality. *J. Comp. Psychol.* **22**, 61–8.

HALL, C. S. (1936*b*). Emotional behavior in the rat. III. The relationship between emotionality and ambulatory activity. *J. Comp. Psychol.* **22**, 345–52.

HALL, J. F., SMITH, K., SCHNITZER, S. B. & HANFORD, P. V. (1953). Elevation of activity level in the rat following transition from ad libitum to restricted feeding. *J. Comp. Physiol. Psychol.* **46**, 429–33.

HARLOW, H. F. (1953). Motivation as a factor in the acquisition of new responses. In Brown, J. S. *et al.* *Current Theory and Research in Motivation: A Symposium*, pp. 24–49. Lincoln: Univ. Nebraska Press.

HUNT, H. F., & OTIS, L. S. (1953). Conditioned and unconditioned emotional defecation in the rat. *J. Comp. Physiol. Psychol.* **46**, 378–82.

LEBO, D. (1953). A simplified method for measuring emotional defecation in the rat. *Science*, **118**, 352–3.

LINDQUIST, E. F. (1953). *Design and Analysis of Experiments in Psychology and Education.* Cambridge, Mass.: Houghton Mifflin Co.

MONTGOMERY, K. C. (1951). The relation between exploratory behavior and spontaneous alternation in the white rat. *J. Comp. Physiol. Psychol.* **44**, 582–9.

MONTGOMERY, K. C. (1952). Exploratory behavior and its relation to spontaneous alternation in a series of maze exposures. *J. Comp. Physiol. Psychol.* **45**, 50–7.

MONTGOMERY, K. C. (1953). The effect of the hunger and thirst drives upon exploratory behavior. *J. Comp. Physiol. Psychol.* **46**, 315–19.

MORGAN, C. T., & GALAMBOS, R. (1942). Production of audiogenic seizures by tones of low frequency. *Amer. J. Psychol.* **55**, 555–9.

MUNN, N. L. (1950). *Handbook of Psychological Research on the Rat.* London: Harrap and Co.

O'KELLY, L. I. (1940). The validity of defecation as a measure of emotionality in the rat. *J. Gen. Psychol.* **23**, 75–87.

PARKER, M. M. (1939). The interrelationship of six different situations in the measurement of emotionality in the adult albino rat. *Psychol. Bull.* **36**, 564–5. (Abstract).

PASCAL, G. R. & SWENSON, C. (1952). Learning in mentally ill patients under conditions of unusual motivation. *J. Personality*, **21**, 240–9.

PATRICK, J. E. (1931). The effect of emotional stimuli on the activity level of the white rat. *J. Comp. Psychol.* **12**, 357–64.

STONE, C. P. (1932). Wildness and savageness in rats of different strains. In Lashley, K. S. (ed.). *Studies in the Dynamics of Behavior*, pp. 1–55. Chicago: Univ. Press.

STOUFFER, S. A., GUTTMAN, L., SUCHMAN, E. A., LAZARSFELD, P. F., STAR, SHIRLEY A. & CLAUSEN, J. A. (1950). *Studies in Social Psychology in World War II.* Vol. **4**. *Measurement and Prediction.* Princeton, New Jersey: Princeton Univ. Press.

THOMPSON, W. R. (1953). Exploratory behavior as a function of hunger in 'bright' and 'dull' rats. *J. Comp. Physiol. Psychol.* **46**, 323–6.

THOMPSON, W. R. & BINDRA, D. (1952). Motivational and emotional characteristics of 'bright' and 'dull' rats. *Canad. J. Psychol.* **6**, 116–22.

VERPLANCK, W. S. & HAYES, J. R. (1953). Eating and drinking as a function of maintenance schedule. *J. Comp. Physiol. Psychol.* **46**, 327–33.

WARNER, L. H. (1932). An experimental search for the 'conditioned response'. *J. Genet. Psychol.* **41**, 91–115.

WATSON, R. H. J. (1954). The effects of age and experience on the performance of the white rat on a test of emotionality. *Bull. Brit. Psychol. Soc.* **23** (inset), 6 (Abstract).

WEININGER, O. (1954). Physiological damage under emotional stress as a function of early experience. *Science*, **119**, 285–6.

YATES, F. (1939). The recovery of inter-block information in variety trials arranged in three-dimensional lattices. *Ann. Eugen., Lond.*, **9**, 136–56.

YERKES, R. M. (1913). The heredity of savageness and wildness in rats. *J. Anim. Behav.* **3**, 286–96.

(*Manuscript received* 9 *July* 1955)

THE ASSESSMENT OF ANXIETY AS AN INTERVENING VARIABLE IN THE DELINQUENT BEHAVIOUR OF M.D. SUBJECTS. GALVANIC SKIN RESPONSE AND LEG-PERSISTENCE INDICES

By JOHN E. TONG

Rampton Hospital, Retford, Notts

Anxiety is accepted as a stimulus-producing response or learnable drive to which certain delinquent behaviour is conditioned by drive reduction, and the relationship between fear, anxiety and pain resistance is discussed. The leg-persistence test is regarded as a pain-resistance test and is shown to be correlated with galvanic skin response fear reactivity. Persistence scores are shown to vary amongst delinquents according to the hypothetical degree of stress engendered by their previous misbehaviour and current hospital confinement. High anxiety (low persistence) subjects are found to be more unstable than others; age and intelligence to have no bearing upon the scores, but psychotic involvement to alter the distributions markedly. The data are discussed in terms of behaviour theory and the Eysenck typology.

I. INTRODUCTION

Fear is currently accepted to be a learned response to pain as a primary stimulus, and to operate in future conditions as a stimulus-producing response, or learnable drive, to which overt behaviour is conditioned by drive reduction (Miller, 1951). Anxiety is defined as generalized fear so that the theoretical relationship between pain resistance and anxiety is apparent, i.e. the greater the anxiety level, the more marked is the overt response to pain, and the less able is the subject to withstand pain. Hence, a crude measure of anxiety is the length of time that a subject can withstand a painful task. (Vide the reports of Hall & Stride (1954) and Stengel, Oldham & Ehrenberg (1955), and their references, for discussion of pain perception and pain reactivity in abnormal states respectively. Relatively little is reported on pain *resistance*, *per se*.)

Anxiety (in some form) is widely regarded as an operative factor in determining delinquent behaviour, but the level of manifest anxiety is partly determined by the environment of the subject at the time of examination. Hence, one would expect more anxiety in the period between the arrest and the trial, than after the sentence, and so on, but at the purely empirical level this is not necessarily a reason why pain resistance should not be used in the psychological prediction of future delinquency in subjects of known delinquent history. Biochemical, pharmacological and pain perception data in support of the view that pain resistance can be regarded as an index of anxiety in delinquent mental defectives, and evidence showing the relationship between pain resistance and ratings of proneness to relapse and the prevalence of psychosomatic disorders, are given elsewhere (Tong & Pugmire, 1955). The current report is concerned with investigations into (*a*) the relationship of scores for induced fear and pain resistance, and (*b*) the varying distributions of pain resistance scores amongst delinquent subjects.

The general theoretical position is that (*a*) a large proportion of delinquent behaviour is a learned response to stress or anxiety; (*b*) the commission of an illegal or punishable act, although providing temporary anxiety reduction, introduces further anxiety or fear;

and (*c*) the possibility of future delinquency is a source of anxiety for some (or many) delinquents. For the sake of clarity the standard terminology of behaviour theory is used (Hull, 1943, 1952, etc.), with the modifications for 'stimulation' and 'response' suggested by Miller & Dollard (1941, p. 59) to cover human social behaviour. The stress reaction hypothesis of delinquent behaviour differs from that of Maier (1949) in so much that stress and anxiety are regarded as synonymous for this purpose, and as stimulus-producing responses.

II. THE PERSISTENCE TEST

The following results were obtained from the use of a Leg Persistence Test similar to that described by Eysenck (1947, 1952), Brady (1948), and Gordon, O'Connor & Tizard (1954). It was applied by the writer as a persistence test in a small-scale study of delinquent mental defectives (Tong, 1954), but following an analysis of subjective reports and specific investigation (Tong & Pugmire, 1955) it was decided that the test is more meaningful regarded as a pain-resistance test. It is well suited to a larger scale study owing to the simplicity of the instructions and the absence of painful connotations. It is also suitable for repeated retesting purposes as there is little or no practice effect, but testing conditions, instructions, etc., must be rigidly controlled.

The test, as used, requires the subject to sit on a chair and hold his right leg horizontally about 2 in. above the seat of a chair of similar height. The score is the time in seconds that elapses before the subject lowers his leg to the chair seat. The instructions are simply—'Hold your leg above the seat of the chair as long as you can'; the maximum score is set at 180 sec.; there is no encouragement or incentive and the patient is given no knowledge of his score. Subjects report a dull pain in the leg after about 20 sec. in the required position (although this point varies greatly), followed by generalized fatigue pain.

III. THE SUBJECTS

These are some 200 male patients at a State Hospital for mental defectives who demonstrate dangerous or violent (delinquent) propensities so that specialized care and supervision is deemed advisable. They were selected randomly from those subjects referred for psychological examination, and are probably weighted slightly with subjects of rather higher I.Q. than the Hospital population as a whole. The disordered behaviour demonstrated by any one subject may have occurred in an institution for mental defectives, in society whilst certified as a mental defective or in society prior to certification.

The main classification of the subjects is into (*a*) those demonstrating no psychiatric or psychological test evidence of possible psychotic involvement, and (*b*) those who demonstrate such evidence. The psychotic subjects are dealt with in a separate section at the end of the paper; I.Q. scores, ages, etc., are dealt with in relevant sections.

IV. LEG PERSISTENCE AND THE GALVANIC SKIN RESPONSE FOR FEAR

The G.S.R. recorder provides a means of observing a subject's reactivity to a fearful stimulus, but difficulty is encountered in applying a satisfactory unit of measurement, particularly when refined statistical methods are required (Woodworth & Schlosberg, 1954, p. 141). Most recent reports refer to the various limitations (Haggard, 1949;

Lacey & Siegel, 1949; Bassett & Ross Ashby, 1954; Venables, 1955), but a method similar to that of Paintal (1951) appeared applicable for the current study. The method requires the establishment of a 'maximum' G.S.R. to which other scores are related; the index obtained is independent of basal ohmic level (Elliott & Singer, 1953), and is speedily obtained.

(i) *Hypothesis*

If the leg-persistence test (as a pain-resistance test) is indicative of the degree of anxiety (or generalized fear), the leg score should be negatively related to the index of galvanic skin responsiveness for induced fear, i.e. low pain score implies a high G.S. fear response.

(ii) *Method*

A record of G.S. responses was obtained by means of a Theratronics Ltd. G.S.R. recorder, recording by means of a galvanometer needle which carries a high voltage and sparks to an earthed recording plate via teledeltos (sensitive) paper. Palmar electrodes were used dry, and gave resistance variations to just beyond 1 megohm.

The leg-persistence test was applied to the patient before attaching the palmar electrodes. The patient was seated in a comfortable chair and the G.S.R. record taken for a resting period of 5 min. to determine the type of reactivity—labile, stable or mixed (Mundy-Castle & McKiever, 1953). Fifteen neutral words (i.e. non-'emotional') were then read out at 15 sec. intervals, and the patient required to give an immediate single-word association response to each stimulus word; the word list was then followed by a rest period when the patient was asked to relax completely and close his eyes. During this final period, three unexpected fear stimuli were given at 30 sec intervals, the first being a brush across the eyelashes with cotton wool, the second a slight touch to the inner lobe of the ear, and the third a sharp 'bang' with a ruler on a piece of card, near to the ear. All deflexions were measured in deflexion units (d.u.), basal level was ignored and the mean score for the neutral words (N.G.S.R.) taken as an index of general basic reactivity; the response showing the greatest deflexion for fear was chosen as the G.S.R. for fear (G.S.R.F.).

(iii) *Results*

The index G.S.R.F./N.G.S.R. × 100 (in d.u.) was taken as an index of relative reactivity to fear and regarded as independent of basal level, temperature, etc. The records of twenty-five randomly selected non-psychotic subjects, excluding labile records and those records showing marked persisting resistance changes during the course of the test, were scored for the index given. The range for the fear index is 250–10,000 and for persistence scores for the same subjects is 180–17; the rank order correlation coefficient (ρ) between the respective scores is $-0{\cdot}576$, which is significant at the 0·01 level of P. It is concluded that the data support the hypothesis as stated, and tend to substantiate the general hypothetical position given in the introduction to the paper.

V. CLASSIFICATION OF THE SUBJECTS

Earlier work (Tong, 1954) merely classified the subjects into various operational groups, i.e. sexual and non-sexual disorders, violent and non-violent disorders, stable and unstable subjects, etc. Following consideration of the theoretical system given earlier, it was decided to apply a classification according to the degree of stress or anxiety expected to be present in a given subject in his present environment, in the light of his previously demonstrated stress reaction pattern, i.e. his form of delinquency. This classification should, hypothetically, result in varying group scores with a test which mirrors the degree of anxiety demonstrated by the subjects, e.g. with a pain-resistance test.

Classifying according to history carries disadvantages with those subjects whose stress reaction pattern has at times been variable in nature. An attempt to classify according to predominant pattern was not successful owing to the difficulty of determining the predominant pattern in most mixed cases. The final classification decided upon for testing against the persistence hypotheses, was as follows:

(*a*) *Homosexual.* Those subjects who have demonstrated sexual misbehaviour with other males, irrespective of other forms of misbehaviour they may also have shown. These subjects were deemed to be in a relatively stressful environment in the Hospital owing to the availability (although very limited) of their stress reaction pattern with its undesirable consequences.

(*b*) *Sexual.* Those subjects who have demonstrated sexual misbehaviour involving females but who have no homosexual histories, irrespective of other forms of misbehaviour they may have shown. These subjects were deemed to be in a relatively stress-free environment owing to their current inability to demonstrate their habitual stress reaction pattern with its stressful consequences.

Sturup (1953) discusses the anxiety shown by Scandinavian sex offenders, particularly the fear of committing further offences, and is able to support his point by giving evidence of the readiness of such offenders to agree to voluntary castration. That anxiety is an important element in the structure of most offenders is a commonplace observation, but the sex offender is faced with additional stress in the way of moral reprobation from the time of his arrest until long after his return to society. The futility of procedures intended to frighten or discomfort anxious subjects into an alteration in their behaviour is evidenced by recent experiments with anxiety as a secondary drive (Gordon & Berlyne, 1954). Tong (1955*a*) found evidence of greater emotional disturbance in sex offenders in prison than was the case for other offenders, and Kay (1954) reported similar results with frustration experiments with young offenders in juvenile institutions. The vitally important aspect of specialized hospital handling of such individuals (as against punishment) is that a more empathetic attitude to their failings prevails for the staff, and for this reason, together with the removal of the fear of further offences, the sex offenders can be expected to adjust well in a protected environment.

(*c*) *Violent.* Those subjects who have not demonstrated homosexual or sexual misbehaviour, but who have shown violent or aggressive traits of a serious nature. For these patients confinement creates daily opportunity for their habitual stress reaction pattern, with the inevitable consequences, and confinement and discipline is deemed to be relatively stressful.

(*d*) *Abscond.* Those subjects who have not demonstrated homosexual, sexual or violent features, but who have repeatedly escaped from Institutions, Approved Schools, Prisons, etc., and committed various forms of crime not covered by the previous headings, e.g. housebreaking, larceny, arson, etc. Institutional confinement is stressful for these patients, almost by definition.

Table 1 gives the numbers of patients falling into these four categories, which are not in hierarchical (hypothetical) stress order, with slight details of their disorders. It will be seen that the homologous offender is fairly common in the sample, and assuming that the disordered behaviour is a conditioned response to stress one must also postulate the existence of a hierarchical habit family of conditioned stress reactions in the case of the subjects with mixed histories. The Abscond group is composed almost entirely of homologous offenders, but this is partly due to the classification system adopted; in this group, however, the offences committed by any one subject are often markedly circumscribed, and with the housebreaker, for example, one finds that a given subject will select similar houses and apply almost identical modes of entry on each occasion. The sex reaction pattern is not so simple, however, and the demonstration of the particular behaviour pattern is not only dependent upon the availability of cues (i.e. the type of female liable to elicit the response, e.g. young girls) but also by optimal hormonal requirements.

Table 1. *Showing the types of offenders falling into the four disorder groups*

Classification according to offences		Total no. of subjects	Total no. of subjects of homologous history	Nos. of subjects who also have histories of offences which are		
				Sexual	Violent	Abscond
Homosexual ($N=30$):	Intercourse	24	2	5	14	10
	Other male indecency	6	3	1	2	1
Sexual ($N=40$):	Sex attack and murder	2	2	—	—	—
	Rape with violence	8	3	—	4	1
	Statutory rape	10	3	—	4	3
	Other heterosex offences	20	11	—	5	4
Violent ($N=40$):	Murder and attempted murder	5	4	—	—	1
	Violent attacks	30	11	—	—	19
	Robbery with violence	5	5	—	—	—
Abscond ($N=40$):	With larceny	35	35	—	—	—
	With arson or damage	5	5	—	—	—
		150				

VI. VARIATIONS IN PERSISTENCE SCORES

(i) *Hypothesis.* Assuming that pain resistance is inversely related to the degree of current anxiety, the groups described earlier should produce differing scores for persistence, dependent upon the degree of stress to which the types of patient are subjected, i.e. relatively high scores for the Sexual group, and lower scores for the Homosexual, Violent and Abscond groups.

(ii) *Results.* Table 2 gives the means and variances of the leg-persistence test scores for the various groups, and the data support the hypothesis as stated. Analysis of variance gives a value for F of 6·954, which is significant at the 0·01 level of P.

Table 2. *Means and variances of the four disorder types classified according to the hypothetical degree of current stress*

		Persistence score	
	N	Mean	Variance
Homosexual group	30	82·94	2421
Sexual group	40	108·32	2280
Violence group	40	76·4	1742
Abscond group	40	66·6	1308

Intergroup F-tests result in significant ratios for comparisons of the variances of the Sexual/Abscond groups ($F=1{\cdot}657$) and Homosexual/Abscond groups ($F=1{\cdot}729$), both significant at the 0·05 level of P. Intergroup t-tests applied where the variance ratios are not significant give the following results:

Sexual *v.* Homosexual	$t=2{\cdot}144$, signif. at 0·05 level
Sexual *v.* Violent	$t=3{\cdot}256$, signif. at 0·01 level
Violent *v.* Abscond	$t=1{\cdot}16$, not significant
Violent *v.* Homosexual	$t=0{\cdot}612$, not significant

Fig. 1 gives the distributions of the scores for the four groups.

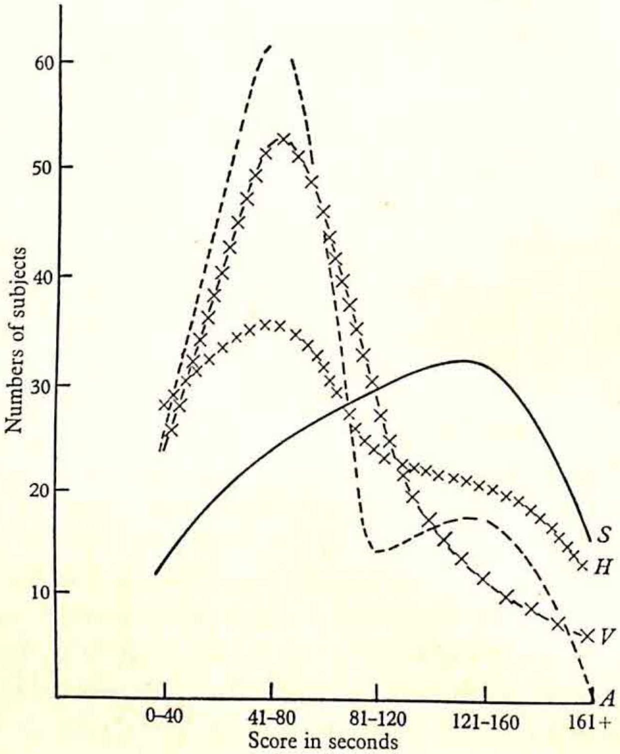


Fig. 1. Showing the distribution of persistence test scores for the four behaviour disorder groups. For each group N is prorated to the equivalent of 120. (S refers to the Sexual group; H, Homosex; V, Violent; and A, Abscond, as defined in the text.)

(iii) *Effect of age*

Age is a factor of some importance in tests involving pain *perception* (Hall & Stride, 1954), and the groups used for the current study give different mean scores for age, in years, as follows: Homosex: mean 29·5, variance 62·4; Sexual: mean 30·7, variance 59·7; Violent: mean 24·2, variance 37·8; Abscond: mean 24·6, variance 38·2. In the earlier study comparing Sexual (including Homosexual) with Non-sexual subjects, the persistence score differences reflected the age differences, and it appeared possible that the persistence score was not therefore a fundamental personality item. The greater differentiation of the current study does not support the view that age is a determinant. The Sexual and Homosexual groups are of similar ages yet give dissimilar persistence scores. Further, rank-order correlations of age and persistence for each of the groups gives the following, not significant, values for ρ: Homosex, 0·02; Sexual, 0·17; Violent, $-0·12$; Abscond, 0·19.

Finally, twenty subjects whose ages were between 15 and 30 years were selected randomly from the Sexual group, and a further twenty subjects of the same age range selected from the Abscond group, i.e. the two extreme groups. The mean persistence score for these two groups is 107·4 and 62·8 respectively; these scores are similar to those given in Table 2, and do not indicate that age determines the differences in score in that table.

(iv) *Effect of intelligence*

Gordon *et al.* (1954) report a low but positive correlation coefficient for I.Q. and leg-persistence score with twenty-seven imbeciles, but no such relationship was apparent in a study of higher grade defectives (Tong, 1954).

In the interpretation of studies involving mental defectives above imbecile level, care must be taken not to confuse the medical/legal notion of mental defect with the psychological/statistical notion of low intelligence. Particularly where behaviour disorders are concerned, the term 'feebleminded' does not necessarily, in practice, imply low intelligence (Memorandum, 1947; cf. Woodward, 1955, and Milner, 1949).

The Vocabulary score of the Wechsler scale (1944) carries the research disadvantages of the full score to a lesser degree and in fact correlates highly with the full I.Q. (Pichot, 1949). It is applied here as a measure of the attainment level, and the mean scores and variances of the four groups are as follows: Sex: mean 13·47, variance 25·2; Homosexual: 16·5, 24·8; Violent: 12·75, 15·2; Abscond: 14·0, 20·2. Rank-order correlations (ρ) between persistence score and vocabulary score for each group are: Homosex, 0·275; Sexual, $-0·406$; Violent, 0·075; Abscond, 0·203. It is suggested that the negative correlation shown by the Sexual group (significant at the 0·05 level of P) is an artifact of the sample selection.

Raven Matrices test scores were used as indices of intellectual potential for twenty-five randomly selected subjects comprising four Homosex, four Sex, ten Violent and seven Abscond. The range for the Matrices score was 8–47, mean 26·16, variance 154·9. A rank-order correlation between Matrices score and Persistence gave a value for ρ of $-0·074$. It is concluded that neither intellectual attainment nor ability are determinants of the persistence score.

VII. HOSPITAL STABILITY

(i) *Hypothesis.* Those subjects who find the Hospital environment to be stressful (irrespective of disorder type) will react more obviously and will therefore be considered to be more unstable and will require more nursing attention than is the case for those who adjust well to the protected environment.

(ii) *Method.* (*a*) The leg-persistence test was applied as an index of current anxiety.

(*b*) The names of the 150 subjects discussed above were listed and the lists handed to three senior nursing staff of a ward, henceforth referred to as the judges. Independently, each judge was asked to select those subjects he knew sufficiently well to make a stability assessment on the basis of behaviour in the ward. The names of the selected subjects known to all the judges ($N=97$) were then listed and the judges asked to rate, independently, each subject as 'stable' or 'unstable'.

(iii) *Results.* Forty-eight subjects were rated in the same direction by all three judges, twenty-eight being 'unstable' and twenty being 'stable'. The other forty-nine subjects were rated 'stable' by two judges and 'unstable' by one judge ($N=23$), or 'unstable' by two judges and 'stable' by one judge ($N=26$).

The mean persistence scores and variances for these four groups are given in Table 3. It will be seen that the means are in descending order corresponding to the degree of certainty (concerning stability) shown by the three judges. Analysis of variance results in a value for F of 4·731 which is significant at the 0·01 level of P (Kempthorne, 1952), and supports the hypothesis as stated.

Table 3. *Means and variances of the persistence scores of the four groups determined by the stability ratings*

Group	N	Mean	Variance
Rated stable by 3 judges	20	105·45	2089
Rated stable by 2 judges	23	90·24	2087
Rated unstable by 2 judges	26	86·14	2354
Rated unstable by 3 judges	28	58·56	1266

Such results, of course, may have reflected a tendency on the part of the judges to select sex offenders for the 'stable' group, as these patients have been shown earlier to be of higher persistence than the other groups. There is a slight tendency in this direction as is indicated by a χ^2 analysis, for which the contingency table is given below. The value for χ^2 is 3·34, which is not significant:

	Stable	Unstable	Total
Sex offenders	14	9	23
Other offenders	29	45	74
	43	54	97

(iv) *Intra-group stability rating.* An important theoretical and practical point is to determine the relationship of persistence and stability for each disorder type. In other words, does a subject's pain resistance score mean the same in stability terms irrespective of the disorder type, or do similar scores carry different meanings for different types of patient? Research into this point is proceeding with more specific tests, but there is evidence to show, for example, with the leg-persistence test that a score below a certain point—about 60 sec.—corresponds with instability and poor prognosis in the vast majority of cases, but for higher scores different confidence limits must be set up for the different types of disorder.

The ninety-seven patients rated for stability can be classified according to disorder type, although the numbers are small, and cross-classified according to whether they were rated as 'stable' (by two or three judges) or 'unstable' (by two or three judges). The mean persistence scores are as follows:

Stable: Homosex 86·6; Sex 141·4; Violent 89·0; Abscond 70·15.
Unstable: Homosex 63·37; Sex 108·2; Violent 69·7; Abscond 59·9.

The numbers of the subjects falling into each group are respectively: Stable: Homosex 10; Sex 14; Violent 6; Abscond 13; and Unstable: Homosex 16; Sex 9; Violent 16; Abscond 13. For this type of study the numbers are too small to be worthy of detailed statistical analysis or interpretation. It will be seen, however, that for each disorder group the Unstable score is lower than the Stable score, but, in addition, the Unstable persistence score for the Sexual group is higher than the Stable score of any other group.

VIII. PERSISTENCE SCORE AND SUBJECTIVE RATING OF FATIGUE/PAIN

Interesting results concerning the patient's own assessment of his fatigue point compared with his objective fatigue point are reported by Hemphill, Hall & Crookes (1952) in respect of anxiety states, hysterics and depressives. In this research a standard finger ergograph was used as the fatigue task and the patient was asked to indicate at which stage he felt tired and then to carry on as long as possible. This technique was found to be inapplicable to the subjects under discussion in the current paper, and consequently an assessment of the patient's subjective opinion of his fatigue/pain level was obtained by asking the question 'Do you ever feel tired or aching, or complain of being tired or aching?' before applying the leg-persistence test. The patient's answer would be classified under one of two headings: (*a*) complaint of tiredness or aching (an affirmative answer or any qualification), or (*b*) no complaint of tiredness or aching (a definite negative answer).

The hypothesis to be examined is that the leg-persistence score is related to a general awareness of fatigue/pain in other, non-specific, situations. The persistence scores of thirty-four patients who gave answers falling under type (*a*) above were compared with the scores of thirty-four patients whose answers indicated no awareness of fatigue/pain, type (*b*). A 2×2 χ^2 test was applied to the data with the persistence score reduced to a dichotomous variable at the median. The value of χ^2 was found to be 4·974, Yates's correction 4·026, which is significant at the 0·05 level of *P*. The actual contingency table is as follows, and indicates that there is a relationship with the persistence scores and general non specific awareness of pain and fatigue:

	Persistence score		
	Below 91	Above 90	Total
'Complaints of tiredness or aching group	25	9	34
'No complaint of tiredness, etc.'	16	18	34
	41	27	68

IX. PSYCHOTIC INVOLVEMENT

The problem of the offender who demonstrates psychotic dysfunction is relatively rare, and even in specialized hospitals the numbers of subjects falling into the various diagnostic categories is usually insufficient to warrant sound conclusions from group studies, particularly if it is desired to cross-classify according to behaviour disorder type. It is generally accepted that the possibility of psychotic involvement must be explored in the early stages of a psychological examination.

Leg-persistence test scores for fifty-eight subjects demonstrating psychological test evidence or psychiatric evidence of psychosis are available. The psychological tests used were Klopfer & Kelley's (1946) system for the Rorschach material, the Wechsler Bellevue scale and the Wechsler Memory scale, the Barrow Hospital Sorting Test (modified) (Hall, 1951, 1952; Tong, 1955*b*), and various conceptual and perceptual tasks. The diagnoses, or diagnostic indicators on tests, suggest: schizophrenia, 28 subjects; post-encephalitis lethargica syndrome, 13; epilepsy, 9; intracranial organic pathological state (undifferentiated), 4; paranoides dementia, 3; manic depressive, 1. Classification according to behaviour shows 9 Homosex; 12 Sexual; and 37 Violent. The mean persistence scores for the behaviour groups are Homosex 35·0; Sexual 69·9; Violent 105·5. It will be seen that the relative position of the three groups is different from that for the non-psychotic

BUREAU OF EDNL. & PSYL. RESEARCH
27.8.70
LIBRARY
J 230

subjects given in Table 2. However, inspection of cross-classification tables reveals that, for these subjects at least, the nature of the psychosis may be the determinant of the persistence score. Of the thirteen post-encephalitic subjects only one scores above 90, whereas only one of the nine epileptics scores below 100. Further for these two groups, eight of the nine Homosex are post-encephalitic, and all the epileptics fall into the Violent behaviour group.

Of the thirty-eight schizophrenic subjects, one is classifiable as Homosex, 9 as Sexual, and 37 Violent, with mean scores of 72·2 and 109·3 for the two latter groups—again a reversal of the position for the non-psychotic subjects. The distribution of the scores for the schizophrenia group is markedly bimodal, with the peaks at below score 50 and above score 180.

The problem of pain resistance in psychotic states may not be one of learning theory but of neural facilitation, but the data at least support the view that early differentiation in diagnosis is advisable, and that any technological conclusions drawn from the early figures for non-psychotics do not necessarily apply where there is gross dysfunction.

X. Discussion

The data given in the paper tend to support the loose extension of academic theory forming the introduction, but in addition, and of equal importance, open up many problems which are approachable by systematic study. At this stage, allowing for the validity of the theory, discussion concerning the respective values of molar or molecular levels is of slight utility only, but it is suggested that further research will require molecular structure to account satisfactorily for the position of 'anxiety', 'fear' and 'pain'. Knopfelmacher (1953*a*, *b*) and Jones (1954) found difficulty concerning the role of 'stress' or 'frustration' as determined by Maier (1949) and suggested the need for physiological derivation after the lines of Cannon (1932) and, later, Selye (1950). 'Anxiety' is a concept similar to 'stress' and may be better contained in a theoretical system containing physiological variables, possibly similar to that of Hebb (1949).

Concerning the role of anxiety, the experimental data relating the scores of the G.S.R. for fear and the persistence scores are of interest in the light of Eysenck's theoretical model, particularly a recent paper (1955). In the latter article Eysenck reviews the anxiety (as a secondary drive) experiments of Taylor (1951) and Spence & Taylor (1951), etc., where the experimental situation involved the use of the M.M.P.I. derivative—the Taylor Anxiety Scale. The point is made that this test correlates highly with the Maudesley Medical Questionnaire which is highly loaded with the Neuroticism factor of Eysenck's model, and it is suggested that the anxiety indices are therefore measures of Neuroticism. Leg-persistence test data are used by Eysenck (1947, 1952) for analysis of the typology, and the question arises whether the test data are better regarded as indices of fear or anxiety in the behaviour theory structure rather than in the typological sense that they have been applied.

The same paper by Eysenck (1955) raises the question of the 'psychopath' in his framework. The subjects discussed earlier certainly fit the description of 'psychopath' given by Scott (1954). Eysenck's typology associates, to some extent, psychopaths, criminals and hysterics, and he quotes their characteristics of 'lying, stealing, sexual delinquency, avoidance of responsibility, etc.' The data given earlier, taken at the empirical level alone, at least suggest that the group 'psychopath' or 'criminal' is probably not unitary but spread along the dysthymic-hysteric continuum, possibly in relation to the type of behaviour disorder. It is important to note concerning researches comparing anxiety states, hysteria and psychopathy, when psychiatric diagnoses are involved, that the terms are at two different levels of semantics. Anxiety state and hysteria are usually (in the psychiatric sense) based upon the intervening variables of some form of analytical type theory, whereas psychopathy is usually descriptive of behaviour alone.

The general problem of pain resistance in theoretical structure is discussed elsewhere (Tong & Pugmire, 1955) at a molecular level, but in the comparison of studies involving cutaneous pain (Hall, 1953) and dull pain such as that of muscular tension, the duality of pain must be considered (Morgan & Stellar, 1950).

For delinquency theory the main points need corroboration with non M.D. subjects, but the following features appear fairly clear:

(*a*) Due consideration should be made of the environment of the subject at the time of the examination. For example, is a comparison of free normals with imprisoned offenders a valid study for the differentiation of personality variables by means of a test (Porteus, 1954)?

(*b*) Statistical studies which regard offenders as forming a homogeneous group should be interpreted with care as the nature of the offence may be an important variable.

(*c*) The psychotic offender may require separate consideration with any technique involving prediction of stability.

(*d*) The importance of anxiety is emphasized by experiment, and for therapeutic purposes anxiety may be better regarded as fundamental rather than symptomatic. Following other researches with anxiety, punitive measures are unlikely to succeed in extinguishing unwanted behaviour, at least with a large proportion of subjects of the type discussed.

The writer wishes to thank Dr G. W. Mackay, Medical Superintendent of Rampton Hospital, for his interest and discussions, and for permission to publish reference to his case material. Thanks are also due to Dr S. L. Pugmire and other medical staff for reference to case material and discussion, and to Messrs Edwards, Taylor and Tuddenham (nursing staff) for their stability ratings, and to Prof. Drew (Bristol) and Mr R. Edwards (Sheffield) for comments and discussion at various stages in the study. The Theratronics Ltd. G.S.R. recorder was constructed as a modification of the circuit supplied by Mr J. Withers of Maudesley Hospital.

REFERENCES

BASSETT, M. & ROSS ASHBY, W. (1954). The effect of electro-convulsive therapy on the psycho-galvanic response. *J. Ment. Sci.* **100**, 632.

BRADY, M. (1948). Suggestibility and persistence in epileptics and mental defectives. *J. Ment. Sci.* **94**, 441.

CANNON, W. B. (1932). *The Wisdom of the Body*. New York.

ELLIOTT, D. N. & SINGER, E. G. (1953). The Paintal index as an indicator of skin resistance changes to emotional stimuli. *J. Exp. Psychol.* **45**, 429.

EYSENCK, H. J. (1947). *Dimensions of Personality*. London. Routledge and Kegan Paul.

EYSENCK, H. J. (1952). *The Scientific Study of Personality*. London: Routledge and Kegan Paul.

EYSENCK, H. J. (1955). A dynamic theory of anxiety and hysteria. *J. Ment. Sci.* **101**, 28.

Gordon, S., O'Connor, N. & Tizard, J. (1954). Some effects of incentives on the performance of imbeciles. *Brit. J. Psychol.* **45**, 277.

Gordon, W. M. & Berlyne, D. E. (1954). Drive level and flexibility in paired-associate nonsense-syllable learning. *Quart. J. Exp. Psychol.* **6**, 181.

Haggard, E. A. (1949). On the application of the analysis of variance to G.S.R. data. I. The selection of an appropriate measure. *J. Exp. Psychol.* **39**, 378.

Hall, K. R. L. (1951). The testing of abstraction with special reference to impairment in schizophrenia. *Brit. J. Med. Psychol.* **24**, 118.

Hall, K. R. L. (1952). Conceptual impairment in depressives and organic patients of the presenile age group. *J. Ment. Sci.* **98**, 256.

Hall, K. R. L. (1953). Studies of cutaneous pain: a survey of research since 1940. *Brit. J. Psychol.* **44**, 279.

Hall, K. R. L. & Stride, E. (1954). The varying response to pain in psychiatric disorders; a study in abnormal psychology. *Brit. J. Med. Psychol.* **27**, 48.

Hebb, D. O. (1949). *The Organization of Behavior.* New York: Wiley.

Hemphill, R. E., Hall, K. R. L. & Crookes, T. G. (1952). Fatigue and pain tolerance in depressive and psychoneurotic patients. *J. Ment. Sci.* **98**, 433.

Hull, C. L. (1943). *Principles of Behavior.* New York: D. Appleton-Century.

Hull, C. L. (1952). *A Behavior System.* New Haven.

Jones, L. C. T .(1954). Frustration and stereotyped behaviour in human subjects. *Quart. J. Exp. Psychol.* **6**, 12.

Kay, B. (1954). Reactions of delinquents and other groups to experimentally produced frustration. *Brit. J. Delinq.* **4**, 245.

Kempthorne, O. (1952). *Design and Analysis of Experiments.* New York: Wiley.

Klopfer, B. & Kelley, D. M. (1946). *The Rorschach Technique.* London: Harrap and Co. Ltd.

Knopfelmacher, F. (1953*a*). Fixations, position stereotypes and their relation to the degree and pattern of stress. Part 1. *Quart. J. Exp. Psychol.* **5**, 108.

Knopfelmacher, F. (1953*b*). Fixations, position sterotypes and their relation to the degree and pattern of stress. Part 2. *Quart. J. Exp. Psychol.* **5**, 150.

Lacey, O. L. & Siegel, P. S. (1949). An analysis of the unit of measurement of the G.S.R. *J. Exp. Psychol.* **39**, 122.

Maier, N. F. R. (1949). *Frustration: A Study of Behaviour without a Goal.* London: McGraw Hill.

Memorandum (1947). *Interpretations on Definitions in Mental Deficiency Act*, 1927. (B.M.A. and Magistrates Association, Joint Committee). London: B.M.A. House.

Miller, N. E. (1951). Learnable drives and rewards. In *Handbook of Experimental Psychology*, ed. S. S. Stevens. New York: Wiley.

Miller, N. E. & Dollard, J. (1941). *Social Learning and Imitation.* New Haven: Yale University Press.

Milner, K. O. (1949). Delinquent types of mentally defective persons. *J. Ment. Sci.* **95**, 842.

Morgan, C. T. & Stellar, E. (1950). *Physiological Psychology.* New York: McGraw Hill.

Mundy-Castle, A. C. & McKiever, B. I. (1953). The psycho-physiological significance of the G.S.R. *J. Exp. Psychol.* **46**, 15.

Paintal, A. S. (1951). A comparison of the galvanic skin responses of normals and psychotics. *J. Exp. Psychol.* **41**, 425.

Pichot, P. (1949). *Les Tests Mentaux en Psychiatrie.* Paris: Presses Universitaires de France.

Porteus, S. D. (1954). Maze test qualitative aspects. *Brit. J. Med. Psychol.* **27**, 72.

Selye, H. (1950). Adaptive reactions to stress. Quoted by Jones (1954).

Scott, P. (1954). Criminal Justice Act of 1948 and the psychiatrist. *Howard J.* **9**, 20.

Spence, K. W. & Taylor, J. A. (1951). Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. Exp. Psychol.* **42**, 183.

Stengel, E., Oldham, A. J. & Ehrenberg, A. S. C. (1955). Reactions to pain in various abnormal mental states. *J. Ment. Sci.* **101**, 52.

Sturup, G. K. (1953). Sexual offenders and their treatment in Denmark and other Scandinavian countries. *Int. Rev. Crim. Policy*, **4**, 1. United Nations ST/SOA/Ser./M. 4.

Taylor, J. A. (1951). The relationship of anxiety to the conditioned eyelid response. *J. Exp. Psychol.* **41**, 81.

Tong, J. E. (1954). Variations in persistence scores with M.D. behaviour disorders. Rampton Hospital (not published).

TONG, J. E. (1955*a*). A preliminary note on the perceptual (Rorschach) characteristics of the sexual offender. *Brit. J. Delinq.* 5, 304.

TONG, J. E. (1955*b*). Abstraction ability and impairment. The Barrow Hospital Sorting Test in clinical psychological testing. *Brit. J. Med. Psychol.* 28, 19.

TONG, J. E. & PUGMIRE, S. L. (1955). An experimental examination of prognostic and chemotherapeutic aspects of fatigue and pain tolerance in disordered mental defectives. To be published.

VENABLES, P. H. (1955). The relationship between P. G. R. scores and temperature and humidity. *Quart. J. Exp. Psychol.* 7, 12.

WECHSLER, D. (1944). *The Measurement of Adult Intelligence.* Baltimore: Williams and Wilkins.

WOODWARD, M. (1955). The role of low intelligence in delinquency. *Brit. J. Delinq.* 5, 281.

WOODWORTH, R. S. & SCHLOSBERG, H. (1954). *Experimental Psychology.* London: Methuen.

(*Manuscript received* 7 *August* 1955)

Bureau Ednl. Psy. Research
DAVID H TRAINING COLLEGE
Dated ....
Accs No ....

NEUROTICISM IN CHILDREN

I. AN INVESTIGATION OF NORMAL AND NEUROTIC GROUP DIFFERENCES

By J. G. THORPE* and D. P. JAMES

Institute of Psychiatry (Maudsley Hospital), London

Three groups of children aged 11–15 years, equated on sex, age and intelligence, were given a battery of neuroticism questionnaires and objective performance tests. The experiment was designed to throw light on the effects of the testing situation upon test scores. The results indicate that all the tests, questionnaire or objective performance alike, are to some extent affected by the testing situation, the questionnaires being so much affected as to be virtually useless in discriminating between normal and neurotic groups when the testing situation is more strictly controlled. The objective performance tests still appear to behave in the manner expected of them in spite of the additional controls.

I. Introduction and problem

Early work on the dimensional analysis of personality by Eysenck (1947, 1952) and his colleagues has indicated the existence of at least three dimensions which are relatively independent of one another. Besides intelligence they are—Neuroticism, Introversion-Extraversion, and Psychoticism. It is with Neuroticism that the present experiment is concerned.

Work with adult neurotics suggests that in many cases these patients have a long history of emotional disorders reaching back into their childhood. Would it be possible, therefore, to identify and measure the same factors of neuroticism or emotional instability in children as well as adults, and would the same types of tests be suitable for this measurement? An early attempt to provide answers to these questions was made by Himmelweit & Petrie (1951); the outcome of their work was strongly positive. Even more positive were the results reported by Connor (1952) who used many of the measures originally designed by Himmelweit & Petrie, and whose work was carried out under the guidance of Prof. P. E. Vernon. Similarly, Eysenck & Prell (1951) arrived at a favourable conclusion. It seemed worth while to repeat these studies on other groups and also to introduce certain controls into the experiment which had not been employed before, but which appeared essential before any definitive conclusions could be reached.

The main innovation in the area of additional controls was connected with the problem of eliminating the possible effects of certain factors of set and attitude which might influence results in a variety of ways. The main experimental design consisted of comparing the test scores of normal and neurotic children, equated on other relevant variables. Predictions were made in each case as to the expected direction of the scores in line with theories derived from work with older subjects. Highly significant differences, such as those obtained by Himmelweit & Petrie, by Connor, and by Eysenck & Prell might reasonably be interpreted as a reflexion of the greater emotional instability of the group of neurotic children. However, in all preceding investigations, neurotic children had been tested in the clinic, which to them undoubtedly was associated with detailed discussions of their symptoms, with highly emotional reactions, and with attitudes and behaviour patterns of a neurotic character. Normal children had usually been tested in their own schools, but even when tested in the clinic there could be no such disturbing associations for them as there had been for the neurotic children. The possibility, therefore, had to be faced that the differentiation between normal and neurotic children might be an artefact caused by the determining tendencies or 'sets' arising from the place where testing occurred.

---

* Now at Banstead Hospital, Sutton, Surrey.

To overcome this difficulty, a plan was adopted which first arose in discussions with Mr D. Prell. In brief it was decided to equip a large caravan as a mobile laboratory, to take this caravan to the schools to be sampled, and to carry out all testing on the precincts of the schools in question. By a procedure to be explained later, neurotic children and normal controls were selected in such a way that their selection appeared to be due to chance. In this way it was hoped to eliminate the association for the neurotic child of his symptoms and troubles with the testing, there being no apparent connexion between the caravan and the clinic.

A final check on the effects of set on test scores was provided by testing a further group of children, all of whom were receiving treatment at the Maudsley Children's Out-Patient Department, but this time the testing was carried out in the clinic. If the place of testing and its associated attitudes was important, this could be demonstrated by comparing the test scores of this group with those of the previous group of emotionally disturbed children tested in their schools.

II. THE EXPERIMENT

A. *Subjects*

Three groups of children were involved in the experimental design.

The first group comprised thirty-four children who were receiving psychiatric treatment for emotional disturbances. These children were between 11 and 15 years of age, with a mean age of 12 years 6 months. The second group comprised 106 normal children to serve as controls, and a third group was made up of a further twenty-eight children all receiving psychiatric out-patient treatment. All these groups were equated on sex, age, and intelligence, and there were no reasons for suspecting that the first and third groups differed in their degree of maladjustment.

B. *Tests*

The main criterion in the selection of tests was that they had in previous studies been shown to discriminate between normal and neurotic groups. These tests are too well known to warrant detailed descriptions of them here. The verbal tests are identical with those used by Connor (1952), and brief descriptions of them appear in a paper by Thorpe (1955). Adequate descriptions of most of the objective performance tests will be found in the two books by Eysenck (1947, 1952). Three new measures of neuroticism were introduced—the Body Sway number of reversals score, which was a straightforward count of the number of times the subject's movement in the forward-backward plane changed direction during the Body Sway test, and two Reaction Time scores, obtained from the test fully described by Rodnick & Shakow (1940). Five potential measures of psychoticism were also included.

Verbal tests

- Thurstone's Primary Mental Abilities (11–17 years)
- Sentence Completion
- Annoyances
- Ways to be Different
- Medical Questionnaire
- Worries and Anxieties
- Interests
- Sociometric

Performance tests

- Manual Dexterity (sum of M N O and P scores)
- Track Tracer (amount of time spent in error in $\frac{1}{100}$ sec.)
- Leg Persistence (time in seconds)
- Dynamometer Persistence (time in seconds)
- Static Ataxia (sway in inches)
- Body Sway Suggestibility (sway in inches)
- Body Sway—number of reversals
- Autokinetic Movement (in inches)
- Rail Walk (in feet)
- Tapping—maximum tempo (number of taps)
- Reaction Time—regular preparatory interval (in $_{0}$ sec.)
- Reaction Time—irregular preparatory interval (in $\frac{1}{100}$ sec.)
- Luria—left-hand disturbance
- Dark Adaptation (in grades)
- Mirror Drawing (number of errors)

The following tests of psychoticism were also included:

- Three circles (time in seconds)
- Three squares (time in seconds)
- Waves—amplitude (in millimetres)
- Waves—wave length (in millimetres)
- Tapping—normal tempo (number of taps)

C. *Method*

The names of thirty-four children between 11 and 15 years of age were selected from the records of the Children's Departments at the Maudsley and St George's Hospitals. All the children were attending their schools in the normal way, but were considered to be sufficiently disturbed emotionally to need out-patient psychiatric treatment. Care was taken that none of these children suffered from organic complications, possible psychotic disturbances, or complaints of a physical nature. These thirty-four children were, with the permission of the L.C.C. and the school head teachers, tested on the premises of their own schools.

In order to obtain the normal control group, three children were chosen from the school class of each of the thirty-four neurotic children, who most closely resembled him in sex, age and intelligence. This was made possible by administering the intelligence test to the whole class, and scoring it during the subsequent testing session with the verbal tests. In this way a control group of 106 children equated on sex, age and intelligence with the group of thirty-four neurotics was formed. The children were not informed, however, of the selection procedure. They were told that some of their names would be drawn out of a hat and that these pupils would be doing some other tests (the performance tests) individually in the caravan. By following this procedure it was hoped to eliminate the associations for the neurotic child between his symptoms and the testing.

The third group in the experimental design comprised twenty-eight children, all attending the Maudsley Children's Out-Patient Department for emotional difficulties, and equated for sex, age and intelligence with the first group of neurotics. These children were tested individually on both the verbal and the performance tests, but as in most cases a parent was waiting for the child the test battery had to be somewhat reduced.

In summary, three groups of children, equated on sex, age and intelligence, were involved. The first (neurotic) group was tested in the schools, the second (control) group was also tested in the schools, and the third (neurotic) group tested in the clinic. There were no reasons for supposing that the third group was any more or any less neurotic than the first.

III. RESULTS

The means and variances of the test scores for each of the three groups are given in Table 1.

Table 1

Variable	Expected direction of score indicating neuroticism	Neurotic school tested ($N = 34$)		Control school tested ($N = 106$)		Neurotic clinic tested ($N = 28$)		Critical ratio, normal and neurotic school tested*
		Mean	Variance	Mean	Variance	Mean	Variance	
1. Age in years	—	12·500	1·217	12·475	1·366	12·833	1·191	N.S.†
2. Total P.M.A. score	—	94·917	875·415	92·121	981·000	90·032	1421·312	N.S.
3. Sentence completion	High	3·455	4·4	3·623	4·6	5·285	5·1	0·406
4. Annoyances	High	24·079	96·6	25·399	63·5	24·607	82·3	0·712
5. Ways to be different	High	20·992	128·7	25·004	167·2	23·750	143·9	1·787
6. Medical questionnaire	High	8·879	16·3	9·315	17·4	10·500	20·3	0·233
7. Worries and anxieties	High	12·172	35·5	11·872	36·0	13·964	38·6	0·256
8. Interests	High	15·700	36·5	15·445	33·1	14·607	25·9	0·216
9. Sociometric	Low	0·204	0·081	0·312	0·088	N.G.‡	N.G.	2·038
10. Manual dexterity	Low	201·147	559·5	203·264	361·7	196·880	411·6	0·475
11. Track tracer	High	116·895	5078·4	132·973	8485·7	N.G.	N.G.	1·355
12. Leg persistence	Low	30·547	464·5	41·586	787·6	29·793	422·9	2·406
13. Dynamometer persistence	Low	30·517	1374·1	46·808	1918·9	22·596	850·0	2·131
14. Static ataxia	High	0·735	0·451	0·636	0·378	0·570	0·512	0·847
15. Body sway	High	2·183	6·3	2·293	5·9	2·527	6·2	0·224
16. Body sway reversals	High	8·937	132·4	7·896	125·4	N.G.	N.G.	0·463
17. Autokinetic	High	23·844	442·1	20·473	565·7	N.G.	N.G.	0·787
18. Rail walk	Low	97·109	304·0	91·500	476·6	N.G.	N.G.	1·530
19. Tapping maximum	Low	59·006	146·1	56·353	86·0	N.G.	N.G.	1·174
20. Reaction time—regular	High	31·569	88·0	29·863	42·9	33·629	258·2	1·116
21. Reaction time—irregular	High	34·100	59·5	33·360	68·9	36·518	297·8	0·516
22. Luria	High	6·027	47·3	5·649	36·9	N.G.	N.G.	0·286
23. Dark adaptation	High	22·668	20·5	21·799	19·2	N.G.	N.G.	0·984
24. Mirror drawing	High	61·916	7128·4	51·174	5375·9	62·504	1287·6	0·666
25. Three circles	—	16·993	148·4	17·108	83·3	N.G.	N.G.	0·051
26. Three squares	—	24·137	191·9	24·778	157·8	N.G.	N.G.	0·240
27. Waves—amplitude	—	20·698	11·9	19·646	12·6	N.G.	N.G.	1·538
28. Waves—wave-length	—	92·755	387·1	89·418	457·6	N.G.	N.G.	0·842
29. Tapping—normal	—	34·221	199·4	35·767	169·2	N.G.	N.G.	0·566

* A CR of 1·65 or more is required for significance at or beyond the 5% level.

† N.S., not significant.

‡ N.G., not given.

IV. Discussion of results

Discussion may first be centred around the test performance of the neurotic and normal groups tested in the schools. It will be seen from the table that the seven verbal tests (numbered 3 to 9) fare rather badly on the whole in terms of the predictions that were made for them, only three of them behaving as predicted. Moreover, the difference between the groups on the Ways to be Different test would have been statistically acceptable had the *opposite* prediction been made. In view of the results of both Himmelweit and of Connor, in which the same or similar tests were shown to discriminate in the predicted direction at high levels of statistical confidence between neurotic and normal groups, the present finding is rather serious in its import. At the same time a considerable weight of evidence suggests, as is now well known, that tests of this type are notoriously unreliable owing to the effects of the testing situation, the sex of examiner, the motivation of the subjects, and so on. Meehl & Hathaway (1946) are two of the many writers who have drawn attention to this. Faw (1948) found changes in Bernreuter responses according to the testing situation, over 15% of the subjects changing more than fifty percentile points from one situation to another. Eysenck, too, is acutely aware of this difficulty which leads him to place more reliance upon tests of the objective performance type in the investigation of all basic personality dimensions. He writes: '(Performance tests) come nearest to the ideal of objective empirical investigation; they suffer less from the evils of voluntary distortion....Our hope of building up a true science of personality rests on the development of this type of test more than on any other' (1950*a*, p. 214). It is to this group of tests that we must now turn.

In the results for the fifteen objective performance tests of neuroticism administered individually in the caravan, all except four behave in the manner expected of them. These exceptions are the Track Tracer, Body Sway, the Rail Walk, and the Tapping (maximum tempo) tests. Of these the results for the Body Sway test are perhaps the most serious, this test having previously shown good discrimination between groups of normals and neurotics, whether adults or children. A detailed discussion of this result will follow later. Of the eleven individual tests behaving as predicted, only two (the two tests of persistence) show differences between the normal and neurotic groups large enough to be statistically significant.

The implications of the results of comparing the performances of neurotic and normal children tested in their schools are that there is little hope of using group tests of the questionnaire variety in this way for the screening of such groups. Most reliance can at this stage be placed on the sociometric test (Thorpe, 1953) for which a significant tendency for neurotic children to be less popular in terms of this test than the normal controls is demonstrated. For the remainder of the verbal tests, whether considered individually or together, their screening efficiency is virtually zero. Such a conclusion would not, however, obtain for the individual performance tests, eleven of the fifteen showing the differences between the groups which were predicted. Even though only two of these eleven tests showed statistically significant differences, that the overall results are not due to chance is indicated by a χ^2 of 3·3 (1 D.F.) on a hypothesis of chance success.* This value of χ^2 approaches significance at the 5% level and it may be taken to mean that the individual tests on the whole have given results in line with previous findings, even though, in the present case, such factors as the attitudes and 'sets' of the subjects have been thoroughly controlled.

---

* Though this test is not strictly applicable as the tests are not independent, this trend should be compared with that for the verbal tests where less than half of them behaved as predicted.

The psychoticism test performances of the normal and neurotic groups show no startling discrepancies, none of the five tests showing significant group differences.

The value and importance of carrying out a study with the additional control of the factors of 'set' and attitude is further indicated by a consideration and evaluation of the test performances of clinic children in the clinic situation. If the scores of this group are compared with the scores of the first group of clinic children tested away from the hospital, the following facts come to light. First the verbal tests. Of the six available comparisons five indicate a more neurotic score for the clinic-tested group. If the six scores are standardized and combined with equal weights into a 'verbal test score', the difference between the scores of caravan- and clinic-tested children approaches significance on a one-tailed 't' test ($t=1\cdot650$), the clinic-tested group scoring more neurotically than the caravan-tested group. One must conclude, therefore, that testing individually in the clinic tends, on the whole, to evoke more neurotic responses to tests of this type. Of the performance tests eight similar comparisons are available, and of these all but one show a higher neurotic score associated with clinic testing. A combined 'performance-test score' derived in a similar way to the combined 'verbal-test score' shows a difference between the two groups of clinic-tested and caravan-tested children significant at beyond the 5% level ($t=2\cdot282$) in the direction of more neurotic responses being associated with clinic testing. The conclusion would appear to be that the testing situation can and does affect the responses of neurotic children in such a way as to make them appear more neurotic than a similar group tested outside the hospital situation. Not only are the questionnaire responses affected in this manner but also the responses to the objective performance tests. This conclusion points to the need, in future work, for controlling the effects of the testing situation, or at least in some way allowing for them.

It is appropriate at this stage to draw attention to some additional discrepancies between the results of the present study and the results of earlier work. In particular, Himmelweit & Petrie found that the tests which best differentiated their normal and neurotic groups were the tests of static ataxia and suggestibility (a finding also in line with the results reported by Eysenck & Prell). On the other hand, the two tests of leg and dynamometer persistence were inefficient in this respect. In the present study the two tests of persistence are the most efficient, and the tests of static ataxia and suggestibility the least efficient in discriminating between the two groups, even when the clinic-tested group is used for comparison with the normal group in order that the two studies should be comparable. As long as different workers using similar tests on similar groups of patients continue to find such divergent results as these, little faith can be attached to any positive statement about personality dimensions other than that we are, so to speak, scratching at the surface. It has been shown above how the gross differentiation of normal and neurotic groups may be, to some extent, exaggerated by a failure to take into account the testing situation. A further difficulty at present looming large (Eysenck, 1952, p. 153) is consequent upon the disturbing possibility of non-linear regressions obtaining between test scores on the dimension of neuroticism.

Consider, for instance, the following diagram, bearing in mind the discrepancy between Himmelweit's results and the present results noted above.

Let the regressions of the four measures in question on the dimension of neuroticism be as shown. Let us suppose, further, that the normal and neurotic groups of the present study fall at points 'T', and that the groups in the Himmelweit study fall at points 'H'. From the diagram it will be seen that in the present study the persistence tests will discriminate more highly between the groups than will static ataxia and suggestibility; also that in the Himmelweit study suggestibility and static ataxia will discriminate more highly than leg and dynamometer persistence. If the regression of these tests is therefore as shown, and if the four groups of normals and neurotics do lie in the positions indicated, then the results of the two studies are consistent with one another. Two deductions can be made from the diagram. First, that for the two tests of persistence the variance of test scores should be *lower* in neurotic groups than in normal, and that for static ataxia and suggestibility the variance of the scores in the neurotic groups should be *higher* than the corresponding variances in the normal groups. Secondly, if the means for the four groups on each of the four tests are compared, they should lie in the order depicted in the diagram. Regarding the first deduction, a glance at the table will indicate that the deduction is in fact borne out, the neurotics having higher test-score variance on static ataxia and suggestibility than the normals, and lower variance on the two tests of persistence. Regarding the second deduction, there is unfortunately a major difficulty, Himmelweit having used in her study a different dynamometer, and different static ataxia apparatus, and a different test of suggestibility. The test of leg persistence should, however, be comparable for the two studies, the only requirement being that the subject should be seated, and hold out his leg for as long as he possibly can. From the above diagram the group means should be arranged in the following order of magnitude, highest first: present-study normals, Himmelweit's normals, present-study neurotics, Himmelweit's neurotics. In fact the means are as follows:

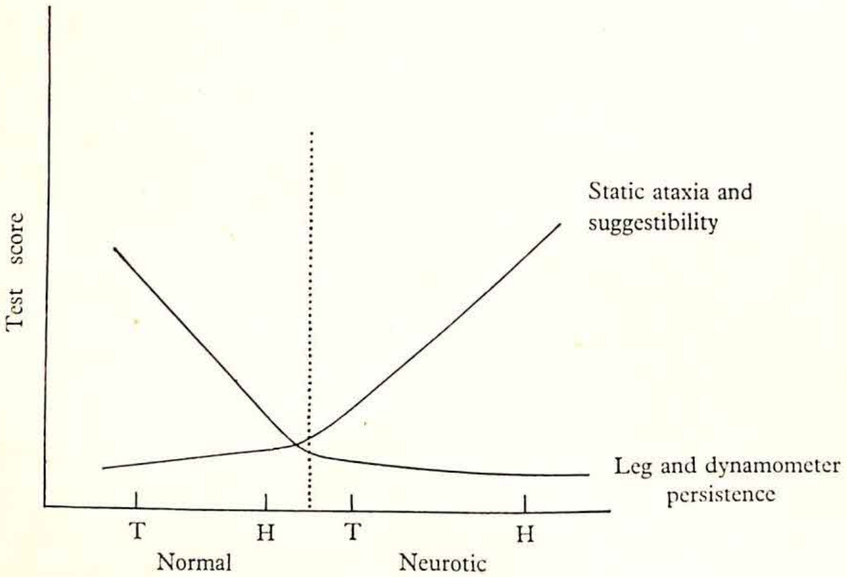

Test score
Static ataxia and suggestibility
Leg and dynamometer persistence
T
H
T
H
Normal
Neurotic

Fig. 1

Present-study normals	41·586	Present-study neurotics	29·793
Himmelweit's normals	34·021	Himmelweit's neurotics	28·024

These figures are in striking agreement with the deduction. It may be concluded, then, that the available evidence suggests that the discrepancies between the results of Himmelweit's study and the present one tend to disappear when account is taken of the types of regression probably obtaining between test scores and the underlying dimension, though more exacting evidence regarding these regression lines is at this stage desirable.

V. CONCLUSIONS

The results of this study point unequivocally to the need for adequate control over such variables as the testing situation and its associated attitudes whenever groups of normal and abnormal subjects are to be compared. It would indeed be of interest and importance to carry out a study with the present battery of tests along the lines of that of Lord (1950), whose work with the Rorschach test strongly suggested the dangers of indiscriminately administering that test. Not only have such controls up to now been omitted from studies concerned with neuroticism in children, but also from the more fundamental studies with adults, upon which the present investigation was primarily based. A recent report by Heron (1955), who demonstrated similar difficulties with the personality tests, would appear to support this contention. Few psychologists would be surprised at such a pronouncement on tests of the questionnaire type. The main import of the present experiment is that even tests of the objective performance variety are to some extent affected in a similar way, though probably to a lesser degree.

The results indicate quite clearly that when the influence of the testing situation upon test scores is taken into account, the questionnaire type of test has no value whatsoever in the discrimination of normal and neurotic children, though this sobering conclusion does not apply to the objective performance tests, most of them showing the sort of group differences that we have come to expect. It should be noted, however, that the obtained differences were smaller than has previously been the case.

A further pitfall in which research workers may indeed find themselves becomes obvious when it is realized that the customary way of identifying factors resulting from factorial analysis is to rotate them into maximum conformity with the discriminatory powers of each test (Eysenck, 1950*b*). As some tests are likely to be more affected by the testing situation than others, it is obvious that the final rotation will, to some extent, be determined by these extraneous factors. It is proposed to follow the present article with a second one which will deal with the factorial analysis of these new data unaffected by these extraneous variables.

There is some evidence in the above report, suggestive though not decisive, that the relation of test scores to neuroticism may be non-linear in some cases, and it has been shown how this may result in apparently startling discrepancies between the results of studies in which even the *same* tests have been applied to similar groups. A test showing good discrimination between two groups in one study may not discriminate at all between two similar groups in another if the only difference between the two studies is that all groups are selected from different points along the neuroticism continuum.

Finally, attention may be drawn to the central difficulty of all research based upon criterion group differentiation. Among adults it has been found that taking a wide range of emotional stability and instability into account, the reliability of psychiatric assessment of degree of instability was no better than 0·5 (Eysenck, 1952); in other words, there is very little agreement among psychiatrists on the nature of the criterion. With children one might reasonably expect the reliability to be considerably lower, the bulk of evidence on the child's difficulties coming from the child's parents or teachers rather than from the child himself. The degree of normality of the normal groups of children is also often open to suspicion for the same reasons. One of the normal children of the present study (now 18 years of age) has recently been admitted to this hospital with a diagnosis of temporal lobe epilepsy with its associated behaviour disturbance. It is hard to see how this difficulty of criterion group selection may be circumvented, but it is clearly a central problem to this type of research.

The writers would like to express their appreciation to Prof. H. J. Eysenck who directed the present study; to the L.C.C. for their kind permission to test the children in their schools; to Drs Himmelweit and Petrie who gave generously of their time to help in the construction and selection of tests, and to the Directors of the Central Research Fund of the Bethlem Royal and Maudsley Hospitals, without whose financial help the study would not have been possible.

REFERENCES

CONNOR, D. V. (1952). The effect of temperamental traits upon the group intelligence test performance of children. Ph.D. Thesis, University of London Library.

EYSENCK, H. J. (1947). *Dimensions of Personality*. London: Kegan Paul.

EYSENCK, H. J. (1950*a*). Personality Tests 1944–1949. Chapter in *Recent Progress in Psychiatry*, vol. II. London: Churchill.

EYSENCK, H. J. (1950*b*). Criterion analysis—an application of the hypothetico-deductive method to factor analysis. *Psychol. Rev.* **57**, 38–53.

EYSENCK, H. J. (1952). *The Scientific Study of Personality*. London: Routledge and Kegan Paul.

EYSENCK, H. J. & PRELL, D. B. (1951). The inheritance of neuroticism: an experimental study. *J. Ment. Sci.* **97**, 441–65.

FAW, V. (1948). Situational variations of neurotic scores measured by the Bernreuter Inventory. *J. Cons. Psychol.* **12**, 225–58.

HERON, A. (1955). The effects of real life motivation on questionnaire responses. *J. Appl. Psychol.* **40**, 65–8.

HIMMELWEIT, H. T. & PETRIE, A. (1951). The measurement of personality in children. *Brit. J. Educ. Psychol.* **21**, 9–29.

LORD, E. E. (1950). Experimentally induced variations in Rorschach performance. *Psychol. Monogr.* **64**, no. 316.

MEEHL, P. E. & HATHAWAY, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. Appl. Psychol.* **30**, 525–64.

RODNICK, E. H. & SHAKOW, D. (1940). Set in the schizophrenic as measured by a composite reaction time index. *Amer. J. Psychiat.* **97**, 214–25.

THORPE, J. G. (1953). A sociometric study of London schoolchildren. Ph.D. Thesis, University of London Library.

THORPE, J. G. (1955). An investigation into some correlates of sociometric status within school classes. *Sociometry*, **18**, 49–62.

(*Manuscript received* 11 *August* 1955)

COGNITIVE INFERENCE IN PERCEPTUAL ACTIVITY

By M. D. VERNON

University of Reading

Whenever the sensory data are scanty or ambiguous, or incongruities occur in the perceptual situation, observers tend to employ processes of inferential thinking to arrive at satisfactory identifications. Such inferential thinking also appears when the observer is called upon to make accurate judgements about events which ordinarily might not be closely observed. In making inferences, observers utilize schematically organized knowledge and tendencies to react which they have acquired, either through experiences which are common to everyone as they grow up, or as the result of particular individual experiences and training. Thus some inferences and judgements are closely similar in different individuals, while others show marked individual differences. Evidence supporting this thesis is given in examples drawn from everyday life and from experimental investigations.

I. Introduction

It seems a matter of some interest and importance to consider a type of cognitive process which is not often studied by psychologists, perhaps because it is hard to define and delimit at all clearly; that is to say, the process of cognitive inference which follows immediate perception. This inferential thinking is probably different in nature from the more complex and highly developed forms of thinking—reasoning, abstract generalization and the like. Brunswik (1950) alluded to the processes which occur in identification through making inferences as 'the implicit logic of perceptual quasi-inference'. These processes differ from the more complex types of thinking in that they are partial, intuitive and even non-veridical. What does occur is a rapid 'hit-or-miss' classification of the immediate percept into some previously established category of experience, or 'schema'. At the moment we know little about the nature or structure of such schemata, nor how nor why any particular schema is utilized for a particular identification. All we can say is that, more often than not, identification is reasonably efficient and successful for all practical purposes.

To illustrate how schematic classification takes place, some examples are given here, together with a description of what seems to occur when it is utilized for identification. These examples demonstrate that the making of inferences is by no means always either accurate or logical. It works well enough in a rough and ready way; but it may lead to mistaken conclusions, which are ignored or overlooked when they are of minor importance from the practical point of view. If they are of major importance, then further perceptual investigation of the situation takes place, and is normally supplemented by more accurate and exhaustive processes of abstraction, generalization and logical reasoning.

II. The distinction between identification and exact observation

It would probably be generally accepted that identifications are determined, not only by the immediate sensory data, and by the expectancies set up in an experimental situation, but also by a more general type of cognitive influence—that constituted by an all-round knowledge of the kinds of experience which occur in particular kinds of situation, and of the kind of behavioural reaction which is most appropriate to them. But there are some situations, and in particular those encountered in laboratory experiments, in which it might be expected that an observer would be completely at a loss, because he would never have encountered just such a situation before. It is true that he is sometimes confused and uncertain, and even that his confusion may persist. But more often he arrives at some identification of the data presented to him which satisfies him. Sometimes these identifications may be very different for different observers. In other cases they may be similar; or may fall into two or more types of similarity. But in all cases they seem to indicate that the observer has made some inference from his sensory impressions in order to 'make sense' of the situation; he thinks he knows what is there, and how he intends to react to it. Thus it may be said that the observer has apprehended the whole situation to his satisfaction. (The notion of 'satisfaction', of knowing how to handle the situation, appears to be more in accordance with the concept of 'functional validity' (see Brunswik, 1955) than that of 'correctness'—that is to say, of correspondence of the perceived with the stimulus object. It is true that the latter may ultimately produce a more successful adjustment to the environment; but it cannot be assumed that it is the natural and invariable goal of the observer's activities.) What he has identified agrees with his previous knowledge and experience of that kind of situation, and he knows how to handle it. But this does not mean necessarily that his identification is veridical. Ordinarily of course he will have little difficulty in using the incoming sensory data to identify particular objects. But sometimes what happens to the observer is less simple; though the complexity may be overlooked if insufficient note is taken of the exact nature of the experience of a single observer, or of the different experiences of different observers. If the sensory data are vague, ambiguous or complex, one observer may say: 'It looks like an X', or, 'It might be an X'. Another observer may say: 'It might be a Y', or, 'It looks like an X, but I think it is probably a Y'—the second part of this statement being based upon his expectations, or upon certain other corroborative data, perhaps supplied by sensory data of another mode. Of course in these situations we are never quite certain that different observers are stimulated by the same sensory pattern. But it is unlikely that this pattern varies to the extent that the reports and descriptions of different observers vary. It seems far more probable that the different observers attach a different significance to the various details of the sensory pattern, and therefore identify it differently. Moreover, the same observer may at one time attach more importance to what he thinks the object *is*; and at another time, to what, momentarily, it *looks like*. Not infrequently an observer will say: 'It looks like an A, but I *think* nevertheless that it is a B'.

This is the kind of difference which occurs, for instance, between the so-called 'sensory' and 'object' judgements of the constancy experiments, which may seem to exist simultaneously when at one and the same moment we perceive things as having their 'real' sizes, and also fitting into the particular angular segment they subtend at the eye. In the 'object' judgement, the one we most commonly use in everyday life, we are concerned mainly with what the object *is*, rather than what it looks like. Thus we wish to know whether it is a shadowed surface, or one with parts in relief; whether it is a blue fabric or a greenish one; whether it is a man standing on the mountainside, or a rock twenty feet high but a long way off. These are instances in which judgement is difficult to make and often at fault;* though there are, of course, innumerable situations in which we make the identification easily and correctly. The 'sensory judgement' is less frequently utilized in everyday life, except when the sensory qualities of the stimulus are markedly in conflict with what we infer to be its identity. For instance, we might notice specifically the brilliant bluish purple of the mountains of Skye seen from a distance; or a sea which looks like molten copper in the rays of the setting sun. These aspects are striking and interesting in themselves, more so than the nature of the particular object to which they appertain. Such percepts are the raw material of artists and poets. The biologist also is trained to perceive the exact appearance of the specimens he examines, rather than to make an immediate and possibly erroneous identification (but cf. p. 43 below).

---

* An instance of this occurred recently to the writer and a friend. An object moving along the sky-line of a hill was identified by the writer as a rabbit, and by the friend as a bird. It proved to be the upper part of a man's body, the lower part of which was hidden by the hillside.

There are occasions, even in everyday life, when we are misled by such unusual sensory impressions to make an incorrect identification because our knowledge and our expectations are inadequate to confirm or contradict them, at least momentarily so. Starting at a shadow seen peripherally may happen easily. Two other examples drawn from the author's experience are: stumbling on a stair which was thought to be only a shadow; and the impression that the railway cutting was sliding down on to the line behind the train, although what was seen was only the induced after effect of relative movement. These mistakes are probably frequent enough, and most of us make them from time to time. Adams (1923) has described a number of similar instances. But usually we ignore or forget them, because they are not corroborated by other aspects of the situation; or because we proceed to check our assumptions by some form of action which shows us that they were non-veridical. But momentarily we identified them in that way, though clearly the sensory data could not in themselves have given rise to that particular identification.

III. AMBIGUOUS IDENTIFICATIONS IN EXPERIMENTS WITH REDUCED SENSORY DATA

Recently a number of similar instances have been demonstrated under experimental conditions of the anomalies in identification of objects which may occur in certain special circumstances, particularly when the sensory data are scanty or ambiguous.

There is a number of demonstrations and experiments by Ames, Hastorff and Ittelson in which the apparent shape, size and distance of objects, seen monocularly with surroundings obscured, appeared to be determined by what the observer believed to be the identity of the objects. Thus Ames (1951) stated that a trapezoidal-shaped window frame rotated about a vertical axis appeared as a rectangular window frame oscillating to and fro. Hastorff (1950) required subjects to adjust the size of a stimulus object, seen monocularly in a dark field, till it appeared equidistant with a standard object seen binocularly. The setting of apparent equidistance was farther away when the variable was given the name of a comparatively large object (envelope, billiard ball) than when named as a smaller object (calling card, ping-pong ball). Indeed, the size of the variable was adjusted until it equated that of the named object seen at the distance of the standard. Ittelson (1951) found that the distance of known objects such as a playing card and a match-box could be judged accurately when they were seen monocularly with surroundings darkened; and half-sized and double-sized playing cards were judged to be double or half their actual distance. When two ink blots were presented, one double the size of the other, the former was usually judged to be about half as far away as the latter. But this judgement was more variable than the judgements of position of familiar objects.

Now the explanation given for these phenomena has sometimes been that what is perceived is determined by the subjects' familiarity with particular objects, rectangular windows, playing cards, and so on. Hochberg & Hochberg (1952) have controverted this hypothesis, and advanced in evidence for their contentions an experiment in which they presented to monocular vision two screens, one having on it the drawing of a boy, the other of a man of exactly the same size. It might be thought that the subject would suppose the latter to be farther away than the former, because he would expect a drawing of a man to be actually larger than that of a boy. In fact, there was no consistent tendency for the man to be located as farther away than the boy. But if two pictures were shown of the same boy, one large and the other small, the latter was judged to be farther away than the former. Thus it seemed that the essential factor was the identification of the two percepts as having been produced by identical stimulus objects of the same size, placed at different distances.

Again, Gogel, Hartman & Harker (1956) presented double- and normal-sized playing cards at different distances, with monocular vision and an otherwise dark field, and compared the estimates of distance for each of the conditions when it was the *first* estimate made by a particular group of subjects. In other words, the subjects were making absolute estimates of distance, unrelated to other estimates made in similar circumstances. The relationship demonstrated by Ittelson did not hold. There was no tendency to see the double-sized card nearer than the normal-sized card. But in subsequent *comparative* judgements, the perceived distance did vary inversely with the retinal size. Thus familiarity with the particular object was not sufficient to define the situation.

In the Ames experiment, it was postulated that the rotating trapezoidal window was perceived as an oscillating rectangular window because one is accustomed to seeing rectangular and not trapezoidal windows. This does not explain why a cross-bar attached to the frame was seen to bend at various points during the movement. Although this may have been necessary to preserve the impression of rectilinearity of the window, it is certainly not in itself a customary experience. Again, Mulholland (1956) has recently shown that the perceived movement of the trapezoidal window may vary quite considerably in character, especially at high speeds of rotation. Sometimes it may appear to rotate, whereas a rectangular window may seem to oscillate; or both may expand and contract.

Now Ittelson, Kilpatrick and Cantril (see, for instance, Kilpatrick (1955) and Toch & Ittelson (1956)), the protagonists of the 'transactionist' approach, consider that perception in such ambiguous situations is a function, not so much of the assumptions made as to the nature of particular objects, but rather of the assumptions based upon a 'weighted average' of previous experiences and of the present stimulus conditions, so that the response is the 'best bet' in the particular circumstances. There are certain situations in which the previous experience of different observers is so similar that they make 'bets' which are approximately the same. But it is clear from the experiments quoted above that different observers often made different 'bets', and that these varied in different conditions in unpredictable ways. This may have been the result of differences in previous experience, or of different weightings of that experience. What seems most probable is that different observers in different circumstances made different inferences as to the nature and conditions of the situation. Thus it is not possible to predict beforehand, without experimental investigation, just what inferences will be made, and what previous experiences will be utilized.

IV. THE FUNCTIONS OF CONFIGURATIONAL CLASSIFICATION

Rather a different light is thrown upon these conclusions by experiments in which the perceived configuration, although its formal characteristics appear quite definite, nevertheless differs radically from the stimulus data. After the configuration has been perceived, some further cognitive process supervenes, and the observer describes or defines the percept in terms of some familiar object or experience, which commonly differs for different observers. Here there appears to be an original perception in terms of some recognizable configuration with characteristic formal properties; further inferences are then made in order to produce a meaningful identification, but it is doubtful whether they greatly influence the original percept.

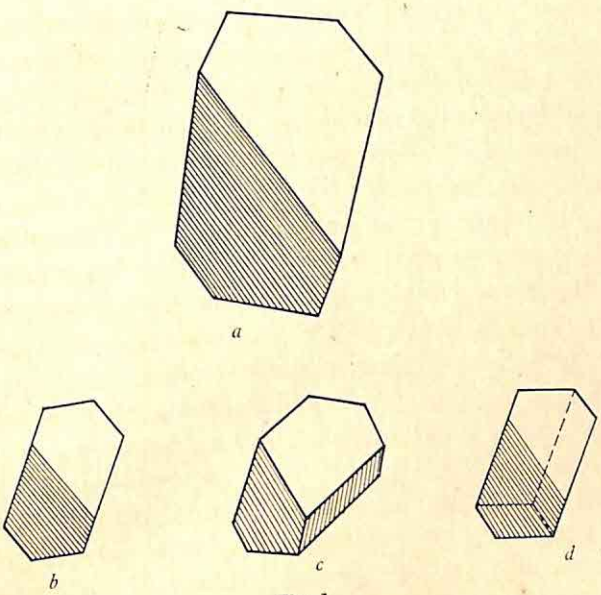


Fig. 1

The first example of these phenomena is given in two sets of experiments by Johansen (1955). In the first, he presented a complex solid made up of two interlacing cubes, in three different spatial positions. The subjects did not know that they were seeing the same object in each of these positions. In the first position they reported seeing a flat shaded surface at right angles to the line of sight; in the second, an object which they called a 'mussel' or a 'lampshade'; and in the third, the two interlacing cubes. In general, they thought they were seeing a new object in each of these positions. It was only after repeated presentation of the three positions in succession that they reasoned out for themselves that it was one and the same object.

In another experiment, observers were shown an irregularly shaped solid (see Fig. 1*a*) at different distances, in binocular vision with daylight illumination. When far distant (about 6 metres), it was seen sometimes as a flat surface, parts of which were shadowed (Fig. 1*b*); and sometimes as a solid object which they likened either to a 'coffin' or the 'model of a barn', or to a 'beam', part of which was shadowed (Fig. 1*c*, *d*). When quite close, it was seen as quite a different object, which was stable but too complex to be named; it could be described only in abstract terms. Furthermore, when they were asked to estimate, by matching, the distance of the object in its first position, their estimates varied according as to whether it was seen as a flat surface, a 'coffin' or a 'beam'.

Hence it appears that the subjects in these experiments first of all perceived the shape characteristics of these objects. They could not identify them exactly in terms of any familiar object—they merely said that they resembled familiar objects. But each perceived configuration was sufficiently definite for them to differentiate it from the other configurations, and to localize it in space. There was no gradual change between one configuration and the next; the point of transition was quite sharp.

Rather similar processes occurred in the movement phenomena described by Michotte (1946) and Johannson (1950). Michotte first demonstrated configurations which consistently had the appearance of mechanical causation, in which the movement of an object was generated by the impact of another moving object. He then presented other configurations which had the appearance of autonomous self-caused movement, in which the expansion and contraction of an object was perceived together with simultaneous translatory movement. This distinction was paralleled in the experiments of Johannson (1950). He presented two black spots moving simultaneously in different ways upon a translucent screen. He found that certain movement configurations were perceived which differed radically from the real movements of the two spots. It was as if in the perception of the movements made they were analysed and re-combined, resulting in the appearance of the new configurations. For instance, two spots moving in opposite directions round circular tracks which touched at one point, were seen as two spots approaching each other along a horizontal line, fusing and separating again; and at the same time the path of their motion rose and fell vertically. Johannson investigated a number of these phenomena, and found that in certain cases, when the movements were referred to a common centre which was at rest, they usually produced the effect of some kind of rather rigid mechanical movement, and were so described by the observers. But in more complex types of configuration, where movement was perceived in relation to a moving centre or frame of reference, it was associated with the more flexible and complex movements of living creatures.

The impression produced by these observations is that quite definite and unambiguous perceptual configurations were perceived, which could not be attributed to the characteristics of the stimuli. They varied to some extent in different observers, and so also did the meaningful identifications given them; but the variations were limited in number. Neither could the perception be said to be a function of identification in terms of some familiar object or meaningful experience. But at the same time the observers spontaneously classified and described their perceptions in terms of such experiences. We may perhaps conclude that, in a non-laboratory situation, an observer would seldom be content with perception of the configuration as such, but would be almost certain to proceed to some process of inference, explanation and classification in terms of familiar experiences, in order to suggest the most appropriate type of behavioural reaction. Nevertheless, the experiments demonstrate the existence of primary schematic classifications in terms of the particular formal characteristics of the configurations.

V. Different types of reaction in constancy judgements

It is clear then that it is not unfamiliarity or inadequacy of the sensory data which in themselves lead to variable and ambiguous identifications. It is also possible to find instances of similar effects in conditions approximating to those of everyday life, for instance in judgements of size and distance associated with the size-constancy phenomenon. We have already noted that constancy judgements may vary according as the observer makes the 'sensory' or the 'object' judgement; that is to say, according to his expectations and intentions in judging the sizes of objects. But these types of judgement are not made with equal ease and spontaneity; and they may vary according to the particular situation in which they are made. Thus Gilinsky (1955) found that with objects of unknown size at great distances (up to 4000 ft.), observers experienced little difficulty in following instructions to make 'retinal' ('sensory') judgements of projected size, and made them fairly accurately. But 'objective' judgements of real size were uncertain and difficult to make, and sometimes involved a considerable amount of reasoning. Joynson (unpublished work) found that with comparatively near objects, the usual spontaneous tendency was to make matches of real size, which were reasonably accurate, by 'allowing for differences of distance'. The observers could also match with fair accuracy according to projected size, when required to do so, by judging the angles subtended by the objects. But unless this method of matching is employed, judgement of projected size is rarely very accurate; according to Piaget & Lambercier (1951) it is more accurate at the age of 7–8 years than in older children and adults. With familiar objects at near distances, the normal method appears to be to make judgements of real size by utilizing the known properties of the objects. Bolles & Bailey (1956) have shown that such judgements can in many cases be made just as accurately without looking at the objects at all.

Thus, in different situations, and with different types of intention on the part of the observer, the size which is immediately perceived may vary. Still other types of appearance may occur in other everyday life situations, for instance when the observer makes a rapid snap judgement about an object of known size, such as a man or a car, at a great distance from him. He may say: 'I think that must be a man although he looks so tiny at that distance.' Or if it is an unknown object, he may say: 'That looks a smallish object, but I know it is a great way off, so it may be a good deal bigger.' Or if he is looking at a photograph of distant mountains, he may say: 'Those don't look very high, but when I saw them myself, they looked much higher.' In other words, apparent size seems to oscillate in some way between values approaching those of real and of projected size. Or both percepts may co-exist, constituting what Piaget (see Michotte, Piaget & Piéron, 1955) has called a 'phenomenal doubling', the one being a quality of what the object is thought to be, the other of what it looks like. However, when the observer is asked to make an accurate quantitative estimate, then he uses his knowledge of the situation to arrive at inferential judgments which tend to fall at one or other of the two polar positions according to his intentions and purpose in making his estimate. These intentions are directed to a greater or less extent by the experimental instructions, but it is impossible for the experimenter to ascertain to what extent. Indeed, it is unlikely that the observer himself can either direct his judgements voluntarily to more than a limited degree, or can be certain how far he has done so. Hence the result of almost all experiments on constancy, that individuals differ from one another in their judgements; and indeed, may be inconsistent with themselves on different occasions, because they have been making different inferences. Whether the actual percepts vary in the same way is impossible to ascertain; but what the observer does with these percepts undoubtedly does vary.

VI. AMBIGUITIES IN THE IDENTIFICATION OF PICTORIAL MATERIAL

A rather different type of inference made from an original percept was studied in an experiment by the author in which the subjects were shown tachistoscopically a series of twelve photographs of scenes and objects of an unusual character, or photographed from an unusual viewpoint; and a series of twelve reproductions of modern paintings. Almost all of these were difficult to identify. From each of the series, three examples were selected which the subject had failed to identify, or had identified incorrectly. These were presented a week later tachistoscopically, together with nine new photographs and nine new pictures. No indication was given the subjects that they would be shown any of the same material again. But it was found that 74 % of the repeated photographs and 64 % of the repeated pictures were recognized immediately and spontaneously as having been seen before. It was not a question of mere general familiarity, but of specific recognition. There were other cases in which the repeated photographs or pictures were reported as being vaguely familiar, or similar to ones seen before; but the experience was obviously a different one.

However, in many of the cases of true recognition, it was the patterns of light and shade and colour that were recognized, and *not* the identification which the subject had made. In other cases, the subject said: 'I remember that, I called it A'; often adding, 'But now I can see it is B', or, 'I can see B as well'. Thus it seemed fairly clear that, when material was difficult to identify with any certainty, the subject often perceived a meaningless visual pattern with sufficient clarity to remember it and discriminate it from other similar material.

But it appeared that the difficulty in identification was not due to the inadequacy of the sensory data. Admittedly these were brief in appearance; but the exposure of $\frac{1}{5}$ sec. was long enough for perceiving simple and unambiguous stimulus material (as was obvious with one or two simpler photographs and pictures). Moreover, the majority of examples were not too vague or dim for discrimination; hence the possibility of subsequent recognition. The difficulty appeared to be either that the objects presented were themselves unfamiliar or were shown in unfamiliar aspects; or that the objects were so incongruous with each other that no consistent hypothesis could explain or reconcile them. It is easy to see how this could occur with examples of modern abstract art. For instance, a composition by Picasso contained a nude torso, a bust and a pattern of leaves, all strongly outlined in black. The latter characteristic gave rise to the frequent response that it was a stained glass window. The nude, though extremely obvious, was omitted by seven of the nine subjects. Other pictures were reported as 'incongruous' or 'not making sense'.

It seems fairly clear then that with this material the subjects were striving to make some meaningful identification of what they perceived, and that inferences were difficult to make because they did not 'fit' one another, or what the subject expected to see,* although some assistance was obtained from the judgement that the pictures were examples of modern abstract painting. Now it is perfectly true that such instances may not occur very frequently in ordinary everyday life. But that they are of considerable importance in situations existing outside the psychological laboratory is demonstrated by recent discussion in the *Lancet* on the difficulties and disagreements between experts in medical diagnosis, and in particular in the reading of X-ray photographs. Johnson (1955) has discussed the differences of diagnosis of chest radiographs made by the same radiographers with the same radiographs after 2 months' interval; and also by American and British radiographers examining the same radiographs. These latter differences were apparently due to the fact that, because the American experts had had much experience of diagnosing pulmonary lesions from very faint shadows, they noticed these more readily than did the British experts. The latter, however, with their extensive experience of pneumoconiosis in coal-miners, more quickly noticed the signs of the changes produced by this disease.

---

* In this case, what the observer would expect to see or would regard as fitting would be pictures which are 'realistic' in accordance with the convention of realism in art which is normally adopted today. As Arnheim (1956) has pointed out, the convention of realism varies from age to age, and present-day 'realistic' pictures are in fact no more exactly representational of the objects they depict than are, for instance, Egyptian sculptures or medieval paintings.

Now Johnson (1953) had previously shown how medical students tended in making drawings of specimens or reports on X-ray photographs to confuse what was really there with what they thought ought to be there. Indeed, some of the students who could not identify a specimen made a better drawing of it than some who did identify it, since the latter tended to make their drawings like the text-book diagrams. It was possible to train the students by discussion and observation to note what was really there rather than what they thought ought to be there. But clearly both the students and the experts were liable to fit their observations to their inferences, and to ignore material which did not fit. In the course of training, the students learnt to be more accurate and objective, to observe appearances rather than to make quick identifications. But the experts had developed new sets of hypotheses (though no doubt of a superior kind) which once more determined their inferences and directed their observations. This controverts to some extent the statements of Gibson & Gibson (1955), that learning automatically and necessarily produces more veridical perception and more accurate discrimination of sensory detail. No doubt the expert radiographers did observe more accurately than did the students; but their observations were still limited by inferences based upon certain acquired hypotheses.

VII. THE DEVELOPMENT OF INFERENTIAL THINKING

Some light is thrown upon the development of inferential thinking in one of its earlier stages by the observations of Piaget & Stettler-von Albertini (1954). They studied the reactions of children from 3 to 4 years upwards when they were shown superimposed simple forms (circle, square, etc.), forms with parts of their outlines distorted or incomplete, and superimposed incomplete forms. It appeared that even quite young children (under 4 years) could perceive simple 'good' forms such as circles and squares directly and immediately. But if these were superimposed, still more if they were mutilated or incomplete, some form of 'perceptual activity' was necessary before the child could reconstitute the simple 'good' forms; that is to say, he had to analyse the stimulus material into its parts, compare these parts, image new parts, and co-ordinate them into a new 'whole'. This activity required at first much trial and error, but it resulted in the building up of perceptual schemata relating to geometrical forms which earlier were perceived passively. The schemata could subsequently be utilized in the perception of these forms in many different settings, for instance, when the stimulus forms were superimposed or incomplete. When this had occurred, the child was then able without further perceptual activity to perceive the completed forms immediately. But before children were sufficiently mature to perform the perceptual activity necessary for completing the forms, they were liable to resort to earlier perceptual schemata, those constituted by certain classes of real objects. They were therefore apt at this stage to perceive incomplete and mutilated forms as representations of real objects.

Now Piaget & Stettler-von Albertini state that this type of perceptual activity develops before the child is able to use intelligent reasoning in making perceptual judgements. Piaget had pointed out earlier (1950) that in perceptual activity the observations made are still liable to syncretism and to ego-centricity, to distortion by reference to the framework and through the influence of particular 'sets'; whereas intelligent judgements are 'de-centred' from the self and from the particular perceptual framework. Nevertheless, these processes are not completely isolated from each other. As the child matures, perceptual activity is performed with a greater degree of intelligence; and the perceptual schemata built up are reinforced and modified by intelligent thinking. Thus as soon as the child ceases to accept the sensory data passively (which must occur much earlier with many real-life situations than with meaningless geometrical forms), he begins to make inferences as he actively directs his observations from one aspect of the environment to another, analysing and comparing what he perceives. At each stage, perceptual data are recorded and organized in schemata, which in turn direct subsequent perceptual activities, until the schemata are themselves modified or even discarded. Finally the developing intelligence will intervene whenever the inferences based upon the schemata are found unsatisfactory, and the individual will try to base his judgements on reason. In other situations, such as making constancy judgements, the child gradually learns to make use of corroborative data, such as are afforded by objects intermediate in position between the standard and the variable (see Lambercier, 1946). It is clear that perceptual activity is also involved here.

Some evidence on the functions of perceptual activity, as modified by the effects of an entirely new set of environmental conditions, may be derived from the experiments of Kohler (1953) on the wearing for prolonged periods of spectacles which reversed the visual field. It was found that voluntarily controlled movements could more readily be adjusted to the new conditions than could habitual movements such as walking and reaching for objects. But these both preceded the adjustment of conscious visual experience. In time, conflicting impressions were ignored, but for a while only partially—that is to say, some parts of the field showed reversal, others did not. Things such as unfamiliar printed words, which had to be considered attentively, still appeared in mirror-reversal, after familiar words had changed to a normal appearance. A critical attitude towards visual phenomena aroused old memories of the earlier appearances of objects. Thus it seems that the perceptual schemata were modified and reformed in piecemeal fashion; that they were most easily re-established in accordance with kinaesthetic data consciously experienced in voluntary non-habitual movement; and that the more complex and unfamiliar the visually perceived situation, the more probable it was that inferential thinking would be utilized to balance previous knowledge against immediate visual impressions.

VII. CONCLUSIONS

From these various observations and experimental results we may conclude that, whatever the observer makes of his original percepts (whether or not he is actually conscious of them), usually appears only in the final identifications he gives to them. These identifications may often depend upon processes of inferential thinking and judgement. We may not be aware that we have made such inferences because they have been rapidly and easily made, and easily checked and verified by overt action. Or we may dismiss them from mind when a satisfactory identification is attained, because they are no longer important. Even in unfamiliar and incongruous situations, we rapidly select the most 'sensible' inferences, and ignore or reject the incongruities. If the latter are too glaring and we cannot ignore them, then we may embark upon a process of explicit reasoning in order to reconcile them. But the inferential process may appear more frequently in experimental situations than in ordinary life, because in the former the observer is called upon to make accurate judgements, often quantitatively accurate, of phenomena to which ordinarily he might pay little attention. But the fact that he so readily resorts to the process of inference argues that he may use it more than he realizes in everyday life. If immediate inference does not lead the observer to satisfactory conclusions, particularly if these should prove to be obviously non-veridical, he will endeavour by means of perceptual activity to examine the field with more care, perhaps utilizing other sensory data by handling and manipulating objects. It is only in the last instance, normally, that he is compelled to resort to intelligent reasoning in order to achieve a satisfactory identification (see, for instance, Johansen's experiments described on p. 39). Often social influences are necessary to bring home to him that his observations do not correspond with those of others. Thus Johnson (1955) has pointed out that disagreements in radiographic diagnoses may be resolved by discussion between radiographers.

Little is definitely known as to the nature and the course of the process of inferential thinking which occurs in making judgements on perceptual data. It is so rapid and facile as to be very difficult to observe introspectively; and, indeed, much of it may take place outside consciousness. Presumably it must employ ideas of fittingness, what events agree and belong together, what makes sense; and, as Piaget and Stettler-von Albertini have shown, these ideas develop gradually as we grow up. When the stimulus conditions give rise to ambiguous or incongruous impressions, when we do not know how to react appropriately, we match and compare these impressions with organized memory material, in order to find some explanation which fits existing phenomena to our ideas and tells us what to do. This memory material is organized along the lines indicated in Bartlett's theory of schemata (1932). But the types of schemata employed are obviously many and various. Sometimes they relate to rather narrowly limited classes of familiar objects, as in the playing card and match-box experiments of Ittelson (see p. 37). Sometimes they have been specially developed by training to cover much more complex perceptual situations, such as those described by Johnson (1953, 1955). But many of the experiments described in this paper have demonstrated the functioning of schemata covering broad categories of phenomena, such as those of geometric form, of related shapes and orientations, of related size and distance, of relative movement, and of spatial orientation. It is probable that within many of these are organized not only visual data, but also much tactile data related to impressions of voluminousness and solidity; and of proprioceptive data, related to position and movement of the body in space. Thus the work of Witkin & Asch (1948), of Werner & Wapner (1952) and of Kohler (1953) has shown that the relation of bodily orientation to the perceived position of objects provides highly labile and complex perceptual data. When measurements of these positions are involved, the inferential judgements show a general schematic determination, but are nevertheless very variable between different individuals.

It must be emphasized that the schemata which are utilized in making inferential judgements in general include response tendencies as well as cognitive data. They give some indication as to the most appropriate form of behaviour in response to the cognized situation. Thus when an observer wishes to utilize his observations in some particular way—to make some particular type of judgement—he spontaneously resorts to the appropriate schema. This classifies and identifies his experience; it suggests hypotheses as to the nature of that experience; and it indicates the manner in which he should act, how he should handle the situation effectively. Furthermore, the consequences of action, and the further perceptual data resulting from action, in turn modify and adjust the schemata in order that they may be more adequate to control perception in future. The 'weighted average' is adjusted, and the best 'bet' altered appropriately.

It is the purpose of this paper to draw attention to the part played by inferential thinking in so many of the judgements that are made in perceptual experiments, and probably also in many of the experiences of everyday life. Whatever the nature of the immediate percept—and that is usually hard to determine—it is the inferential thinking which largely directs the activities involved in making these judgements. It is fashionable nowadays to draw attention to the influence of motivational factors and personality characteristics, particularly in producing non-veridical perceptions and individual differences in perception. But except in exceptional circumstances (for instance, of violent need, when the sensory data are minimal), no evidence has been advanced that the effects of these factors are other than slight. They may possibly accelerate or retard perception; or, as the experiments of Smock (1955) seem to indicate, they may cause an observer subjected to undue stress to jump too quickly to conclusions based on inferences made while sensory data are still inadequate. But the part played by schematic categories of experience (which are, of course, frequently affected by motivation), and the inferences to which they give rise, is more than influential; it is determinative. Organized knowledge of the kind of events that occur in certain situations, and the kind of reactions which are appropriate, permeates and directs our whole present experience. Personality factors, in so far as they are of importance, may stimulate us to acquire certain kinds of knowledge, upon which are based our hypotheses about the situation, and hence the aspects of it which we perceive.

REFERENCES

ADAMS, G. K. (1923). An experimental study of memory color and related phenomena. *Amer. J. Psychol.* **34**, 359.

AMES, A. (1951). Visual perception and the rotating trapezoidal window. *Psychol. Rev. Monogr.* **65**, no. 7.

ARNHEIM, R. (1956). *Art and Visual Perception*. London: Faber and Faber.

BARTLETT, F. C. (1932). *Remembering*. Cambridge University Press.

BOLLES, R. L. & BAILEY, D. E. (1956). Importance of object recognition in size constancy. *J. Exp. Psychol.* **51**, 222.

BRUNSWIK, E. (1950). Remarks on functionalism in perception. *J. Personality*, **18**, 56.

BRUNSWIK, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.* **62**, 193.

GIBSON, J. J. & GIBSON, E. J. (1955). Perceptual learning: Differentiation or enrichment. *Psychol. Rev.* **62**, 32.

GILINSKY, A. S. (1955). The effect of attitude upon the perception of size. *Amer. J. Psychol.* **68**, 173.

GOGEL, W. C., HARTMAN, B. O. & HARKER, G. S. (1956). The retinal size of a familiar object as a determiner of apparent distance. *Rep. Army med. Lab.* no. 235.

HASTORFF, A. H. (1950). The influence of suggestion on the relationship between stimulus size and perceived distance. *J. Psychol.* **29**, 195.

HOCHBERG, C. B. & HOCHBERG, J. E. (1952). Familiar size and perception of depth. *J. Psychol.* **34**, 107.

ITTELSON, W. H. (1951). Size as a cue to distance: static localization. *Amer. J. Psychol.* **64**, 56.

JOHANNSON, G. (1950). *Configurations in Event Perception.* Upsala: Almqvist and Wiksels.

JOHANSEN, M. (1955). An introductory study of voluminal form perception. *Nord. Psykologi's Monogr. Ser.* no. 5.

JOHNSON, M. L. (1953). Seeing's believing. *New Biol.* **15**, 60.

JOHNSON, M. L. (1955). Observer error: its bearing on teaching. *Lancet*, ii, 422.

KILKPATRICK, F. P. (1955). *Recent Transactional Perceptual Research: A Summary.* Final Report on Office of Naval Research Contract N 60 nr 27014.

KOHLER, I. (1953). Rehabituation in perception. *Die Pyramide*, parts 5–7.

LAMBERCIER, M. (1946). La configuration en profondeur dans la constance des grandeurs. *Arch. de Psychol.* **31**, 287.

MICHOTTE, A. (1946). *La Perception de Causalité.* Presses Universitaires de Louvain.

MICHOTTE, A., PIAGET, J. & PIÉRON, H. (1955). *La Perception.* Paris: Presses Universitaires de France.

MULHOLLAND, T. (1956). Motion perceived while viewing rotating stimulus objects. *Amer. J. Psychol.* **69**, 96.

PIAGET, J. (1950). *The Psychology of Intelligence.* London: Kegan Paul.

PIAGET, J. & LAMBERCIER, M. (1951). La comparaison des grandeurs projectives chez l'enfant et chez l'adulte. *Arch. de Psychol.* **33**, 81.

PIAGET, J. & STETTLER-VON ALBERTINI, B. (1954). Observations sur la perception des bonnes formes par actualisation des lignes virtuelles. *Arch. de Psychol.* **34**, 203.

SMOCK, C. D. (1955). The influence of psychological stress on the 'intolerance of ambiguity'. *J. Abnorm. Soc. Psychol.* **50**, 177.

TOCH, H. H. & ITTELSON, W. H. (1956). The role of past experience in apparent movement: A revaluation. *Brit. J. Psychol.* **47**, 195.

WERNER, H. & WAPNER, S. (1952). Towards a general theory of perception. *Psychol. Rev.* **59**, 324.

WITKIN, H. A. & ASCH, S. E. (1948). Studies in space orientation. *J. Exp. Psychol.* **38**, 325, 455, 603 and 762.

(*Manuscript received* 21 *October* 1955)

PERSONALITY AND PERCEPTION: THE CONSTANCY EFFECT AND INTROVERSION

By J. AMOR ARDIS
Department of Mental Health
AND ELIZABETH FRASER
Department of Psychology, University of Aberdeen

An investigation of the relationship between shape constancy and introversion-extraversion was carried out, using as subjects twenty-four first-year Psychology students of Aberdeen University; twelve introverts and twelve extraverts, selected on the basis of the Minnesota TSE scale. Results showed (*a*) that introverts showed lower constancy effect than extraverts and (*b*) that there was a tendency for men to show greater constancy effect than women.

There was also some suggestion of the presence of a learning factor in the perceptual process which also distinguished introverts from extraverts. These results are discussed in the light of previous investigations. Hypotheses, which are shortly to be tested, and which might account for the findings, are advanced.

I. INTRODUCTION: SETTING OF THE EXPERIMENT

In the course of experiments with mescaline, it was noticed that the incidence and intensity of perceptual abnormalities varied with alterations in attention to the environment. On the one hand, so long as a practical, reality-orientated, extraverted attitude was maintained (often with difficulty), the environment would be made to appear approximately normal. On the other hand, when subjects lapsed into a dreamy preoccupied state, they experienced marked perceptual distortions. These included hallucinatory movement, tunnel vision, the appearance of networks superimposed on real objects, and various other distortions which have been reported by many other workers with mescaline. The subject has recently been reviewed by Mayer-Gross (1951).

In our own experience, a regular feature of the distorted perceptions under mescaline was a reduction of the constancy effect—one important aspect of a more general disorganization of perception. For example, one subject, viewing a door somewhat obliquely, remarked that it seemed to be shaped like a wedge and that it required a definite effort to see it as he knew it to be, i.e., rectangular. This particular aspect of the disorganization, reduced constancy, or in Thouless's terminology reduced 'phenomenal regression to the real object', corresponded so closely to the presence of an introverted attitude that it was decided to investigate further the relationship, if any, existing between the constancy effect and the personality trait of introversion in normal individuals.

The role of personality factors in the perceptual process has received considerable attention in recent years. The work of Brunswik, Bruner and Postman, Frenkel-Brunswik, Klein and Schlesinger, and many others in this field, requires no exposition in this article. The area with which we are particularly concerned, the constancy phenomenon, has itself been the subject of a considerable literature. The factors affecting constancy (e.g. the number of cues present, the attitude of the observer, the role of learning, and the developmental aspect of the phenomenon) have all been investigated, and experiments on them have given rise to a number of controversial issues. The question of attitude, which is particularly relevant to our own research, has been studied by Klimpfinger (1933*a*) and Holaday (1933) working under Brunswik. They demonstrated that the adoption of an analytical attitude on the part of the observer results in a marked reduction of both shape and size constancy.

Relatively few researches, however, have been carried out on the more specific question of how persons, who are known to differ on a given personality dimension, behave in a constancy experiment. Thouless (1932) did carry out one such investigation. In the course of previous experiments, investigating individual differences in 'phenomenal regression', he had observed that 'those of introverted or schizothymic temperaments saw much more closely to the stimulus or perspective object* than did the extraverted or cyclothymic. It seemed even possible that this might be a measurable result of the greater interest in outside reality of the cyclothyme and in the subjective world of the schizothyme' (Thouless, 1932). An experiment in which fourteen cyclothymes and fourteen schizothymes (selected on the basis of Kretschmer's descriptions) were tested, gave results which, though they failed to reach significance, suggested nevertheless that schizothymes did experience phenomenal regression less markedly than cyclothymes.

In a later study, Weber (1939), using stereoscopic material, confirmed that extraverts showed greater size constancy than introverts, but reported non-significant results for shape constancy, although the tendency was in the expected direction.

In our experiment we are concerned with shape constancy, and an attempt has been made to control experimentally, or partial out statistically, factors which are known, or thought, to influence the constancy effect. It was predicted that extraverts would show a more marked constancy effect in their perception than introverts.

II. CRITERIA OF INTROVERSION-EXTRAVERSION

For the purposes of this experiment, introverts are regarded as those individuals who have a low score, and extraverts those with a high score on the Thinking dimension of the Minnesota TSE scale of introversion-extraversion devised and standardized by Evans & McConnell (1941). The authors constructed this scale to measure three types of introversion-extraversion, Thinking, Social and Emotional, which were isolated by Guilford in his factor analysis. This analysis yielded evidence that the available I–E tests were not measuring a single dimension of personality, and Evans and McConnell devised three distinct sets of item in an effort to develop relatively independent tests dealing with thinking, social and emotional reactions. The result is a scale providing three measures of I–E, which have been shown by the authors to be almost uncorrelated. It was on the first of these, the Thinking I–E score, that our subjects were selected.

III. SUBJECTS

The subjects were drawn from the first-year class in Psychology in Aberdeen University, and their ages ranged from 17 to 24 years. Out of a group of ninety-one previously rated on the TSE Minnesota Scale, twenty-four students were selected: these comprised the six highest and six lowest scorers on the Thinking dimension among the men, and the six highest and six lowest scorers among the women.

---

* By this Thouless means the object as it appeared optically or in contradistinction to its true nature in objective reality.

IV. APPARATUS AND PROCEDURE

Three stimulus objects were employed; a disk 10 in. in diameter, a square of 10 in. diagonal, and a rod 10 in. long and $\frac{1}{8}$ in. thick. All were of metal and painted white. They were supported by means of a clamp and stand and a rotating horizontal metal bar. They were presented to the subject either vertically—when the entire surface (or in the case of the rod the entire length) was visible, or horizontally, when only the edge of the circle or square (or the end of the rod) was visible, or in one of several intervening positions. The actual angle of inclination to the vertical was measured by a suitably attached protractor.

The subject was provided with a chin rest, adjusted to a comfortable height, and the height of the first stimulus object, the metal disk, was adjusted to eye level so that, when it was in a horizontal position, it was seen by the subject as a straight line and, when vertical, as a complete circle. The subject was also provided with ten white cardboard ellipses, all with major axis 10 in. long and with minor axes ranging from 1 to 10 in., and numbered accordingly.

Instructions were standard for all subjects. It was pointed out that the circle, when tilted, would appear to the subject to be elliptical; he was to choose from the ten ellipses given to him, the one which looked nearest in shape to the tilted circle, and call out its number. He was allowed to give intermediate values when the tilted circle seemed to him to fall between two of his ellipses. When inspecting the cardboard ellipses, he held these at right angles to his line of regard and at arm's length, the stimulus object being 7 ft. away against a black background.

A similar procedure was followed with the square and the rod, the square being presented with its diagonals placed vertically and horizontally. In the case of the rod (which, in the horizontal position, was end on to the subject), he was provided with ten rods ranging from 1 to 10 in. in length from which he had to choose, and for the square, a series of diamond-shaped pieces of cardboard, one diagonal always being 10 in. and the other 1 to 10 in.

V. STATISTICAL TREATMENT OF RESULTS

The two estimates obtained for each presentation of each stimulus object were averaged for each subject and the final results, seven readings for each of three kinds of stimulus object for each of twenty-four subjects (a total of 504 readings), were analysed. Analysis of variance was used, with four main variables: I–E, stimulus object, sex, and angle. Results of this analysis are given in Table 1.

Table 1

Source	*F*-ratio	*P*
Introversion-Extraversion	9·71	0·001
Angle	198·70	0·001
Stimulus	<1	—
Sex	2·96	0·10
Sex × stimulus	1·99	—
I–E × stimulus	1·50	—

The main variable, I–E, gave highly significant results, with an *F*-ratio of 9·71 which with d.f. 1 and 420 gives $P < 0·001$. As predicted, introverts showed markedly less 'phenomenal regression' than extraverts, and this was true for both sexes, and for all three stimulus objects, and at all angles. The elicitation of this result, which was quite clear-cut, was the object of the experiment. Other findings will, however, be briefly mentioned.

Of the other three main variables, angles inevitably accounted for by far the largest proportion of total variance and gave a highly significant F-ratio. This was, of course, expected from the nature of the experimental procedure and is, from the point of view of this experiment, quite unimportant.

The different types of stimulus objects—disk, square and rod—had no noteworthy effect on the degree of phenomenal regression as the F-ratio for 'stimuli' was not significant.

There was a fairly strong tendency for women to exhibit less phenomenal regression (i.e. less perceptual constancy) than men. The F-ratio, however, just failed to reach the 0·05 level of significance.

Of the interactions, not one gave a significant F-ratio. There was a very slight suggestion of a differential reaction to the three stimulus-objects by (*a*) men as opposed to women, and (*b*) introverts as opposed to extraverts. The probability of a chance result is too high, however (P = about 0·20), for any importance to be attached to this result.

VI. DISCUSSION

(i) *Introversion-extraversion*

The significant relationship, which we have found to exist in our subjects, between introversion as a personality trait and reduced constancy in perception of shape, is in general agreement with other workers (Thouless, 1932; Izzet, 1934; Holaday, 1933; Klimpfinger, 1933*a*; Weber, 1939). It is also, we feel, in accord with non-scientific everyday experience. It is somewhat surprising that Sheehan (1938, p. 20, footnote) attached no great weight to the theory that personality might be one of the factors determining constancy.

One explanation of the non-significant results of both Thouless and Izzet may be that they were both ostensibly working with the schizothymic-cyclothymic dimension, subjects being selected on the basis of the authors' assessments. Thouless himself (1932) specifically equates schizothymia-cyclothymia with introversion-extraversion, but we now have reason to doubt that, in their pure form, these dimensions are in any way related. Hence, it is not surprising that his results differ from ours, since he was not dealing with introversion-extraversion in as pure a form. Such correspondence as does exist between his results and ours may be attributed to the fact that, in his selection of schizothymes and cyclothymes, he was influenced by considerations of introversion and extraversion.

Weber's method bears superficial resemblances to ours, but his results are less highly significant. Three possible factors come to mind. First, the instructions given to our subjects were such as to motivate them to perceive phenomenologically rather than in terms of the real object. Though in our experiment extraverts and introverts were presumed to be equally motivated, the latter were probably less rigidly bound to customary perceptual patterns—to perception in terms of the real object. Sheehan (1938) has commented on the apparently greater flexibility in the perceptions of subjects showing low constancy. Weber's subjects, on the other hand, were required to perceive in terms of a supposed real object. It appears to us that such a set is less likely to emphasise differences in personal perceptual patterns.

Secondly, Weber's subjects were American sophomores and, as a body, probably contained fewer really introverted individuals than did our group of Scottish students. Whereas the published norms show that 5% of the American population have scores of less than 85, ten of our group (11%) have scores below this level. On the other hand, at the extraverted end of the scale, no such difference was apparent. This implies that our group spread more widely over the scale; and selection of extremes may have resulted in a more marked contrast between our introverts and extraverts.

Thirdly, for our experiment, we selected as subjects twenty-four extremes out of a total of ninety-one, while Weber chose fifty out of seventy-five. It seems likely, therefore, that our introverts and extraverts were more distinct groups than Weber's and that this helps to account for the more significant results we obtained.

(ii) *The sex difference*

Our results which, though they fail to reach significance, suggest a difference between the sexes, the women showing lower constancy than the men, do not agree with those of Thouless, who found a significant trend in the opposite direction. Methods of selection of subjects, and the fact that, in his experiment, personality variables were not partialled out, may account for the disagreement. Klimpfinger (1933*b*) on the other hand, as we have mentioned, whose subjects ranged in age from 3-year-olds to adults, found that up to the age of 12, boys showed lower constancy, and, thereafter, greater constancy, than girls. There is no indication in her article of the significance or non-significance of these results, and again personality variables were not controlled, but one might say that her results, in the relevant age-groups, agree with ours.

(iii) *The learning factor in perception*

One interesting feature of our results, which just failed to be significant, was that the introverted group, and particularly the male introverts, showed a progressive movement towards the low constancy end of the scale as the experiment proceeded, while the extraverts showed no tendency in any direction. Since the order of presentation of stimuli was not randomized (each subject was shown all positions of the disk, then the square, and finally the rod), it is not possible at present to say whether this tendency is a function of the change of stimulus object or an indication of some learning process at work. An experiment, designed to answer this question, is in progress.

(iv) *Theory*

Before attempting to formulate any hypothesis to account for differences in constancy, it might be profitable to review briefly some of the main conditions and cases where low constancy is known to occur. These include (*a*) low illumination, (*b*) the use of a reduction screen, (*c*) monocular vision and (*d*) a relatively unstructured visual field, giving few cues. In addition, we have noted (*e*) that a phenomenologically directed attitude (as opposed to an object-directed one) results in reduced constancy, and (*f*) that low constancy occurs in introverts. We have also found it (*g*) to be a frequent feature of mescaline intoxication.

We would suggest three possible types of reduced constancy: that due to sensory deprivation, that due to reduced attention to the environment, and that due to a failure in the organizational process. The first calls for little comment; (*a*) to (*d*), mentioned above, are obvious examples.

A reduced attention to the environment appears the most plausible explanation of the reduced constancy we have been discussing in this paper. Though superficially similar to the 'sensory deprivation' type, because of the consequent reduction in amount of information being fed in, reduced constancy due to inattention may be more akin, from the point of view of neurological function, to the third type, presently to be mentioned. Experimental investigation of this second type of reduced constancy might well be carried out using 'latent learning' methods, or by 'attention-saturating' techniques, with particular emphasis on its relationship to speed of learning.

The third type of reduced constancy—organizational failure—is, we tentatively suggest, exemplified in mescaline intoxication. A hypothesis along these lines has already been presented elsewhere (Ardis & Drewery, 1954). Therein, the concepts of Hebb (1949) were applied to certain phenomena of mescaline intoxication. The hypothesis may briefly be summarized thus:

The neuronal basis of perception is taken to be the activation of assemblies and phase-sequences of neurones. Activation, inhibition and modification of these complex spatio-temporal units are regarded as determined by (*a*) the nature of the stimulus and the resultant input to the central nervous system, and (*b*) stored experience in a neuronal form. The latter might conveniently be labelled a 'reality-scheme', analogous to the 'body-scheme' postulated by neurologists, and regarded as a vastly complex—and continually developing—type of phase sequence. The interaction between (*a*) and (*b*) might be expressed either as the shaping—by facilitation or inhibition—of the crude response to (*a*) by (*b*); or as the activation of fractions of (*b*) by (*a*). The distinction is verbal.

It is suggested that the interaction is disorganized by mescaline. One consequence of this may be the emergence of a more primitive, sensory-dominated, perception—one example of which is reduced constancy.

With other effects we are not at present concerned. But it may be suggested that mescaline disorganizes maximally those neuronal units which are the basis of complex, recently learned, responses. This subject may also be viewed from the standpoint of learning theory. A paper (Ardis & Drewery) along these lines is in preparation.

The authors of this article are members of a research group in Aberdeen investigating the model psychoses. They are indebted to Miss Sheila Fraser, M.A., Ed.B., who carried out the testing of subjects. She has presented a report of the experiments in a thesis, in part-fulfilment of the requirements of the Ed.B. degree. They would also like to thank Prof. Rex Knight for helpful comment and criticism.

References

Ardis, J. A. & Drewery, J. (1954). Studies in the model psychoses. Communication to Scottish Branch of Royal Med. Psy. Ass.

Evans, C. & McConnell, T. R. (1941). A new measure of introversion-extraversion. *J. Psychol.* **12**, 111–24.

Hebb, D. O. (1949). *The Organization of Behavior*. New York: Wiley.

Holaday, B. E. (1933). Die Grössenkonstans der Sehdinge bei Variation der inneren und ausseren Wahrnehmungsbedingungen. *Arch. ges. Psychol.* **88**, 418–86.

Izzet, T. (1934). Gewicht und Dichte als Gegenstande der Wahrnehmung. *Arch. ges. Psychol.* **91**, 305–28.

Klimpfinger, S. (1933*a*). Über den Einfluss von intentionaler Einstellung und Übung auf die Gestaltkonstans. *Arch. ges. Psychol.* **88**, 551–98.

Klimpfinger, S. (1933*b*). Die Entwicklung der Gestaltkonstans vom Kind zum Erwachsenen. *Arch. ges. Psychol.* **88**, 599–628.

Mayer-Gross, W. (1951). Experimental psychosis and other mental abnormalities produced by drugs. *Brit. Med. J.* **2**, 317–21.

Sheehan, M. R. (1938). A study of individual consistency in phenomenal constancy. *Arch. Psychol., N.Y.*, **31**, no. 222.

Thouless, R. H. (1932). Individual differences in phenomenal regression. *Brit. J. Psychol.* **22**, 216–41.

Weber, C. O. (1939). The relation of personality trends to degrees of visual constancy correction for size and form. *J. Appl. Psychol.* **23**, 703–8.

(*Manuscript received* 3 *September* 1955)

VISUAL DISCRIMINATION OF ORIENTATION BY OCTOPUS

By N. S. SUTHERLAND

Institute of Experimental Psychology, University of Oxford

Recent research on visual discrimination has largely ignored one of the major problems in this field, namely what are the properties of shapes to which animals are responding when they discriminate them: a solution to this problem must precede attempts to arrive at the mechanism behind shape discrimination. Experiments were conducted on the discrimination of orientation by octopuses. It was shown that discrimination of orientation was easy where the orientations to be discriminated fall in the vertical and horizontal directions, extremely difficult where the orientations are midway between the vertical and horizontal directions. Various transfer and memory tests were carried out which confirm that the animals were discriminating orientation and not other properties of the figures. A tentative explanation of the results is proposed. An analysis of serial effects is attempted which suggests a possible interpretation of some of Boycott & Young's findings on the effects of lesions to the verticalis lobe in octopus.

INTRODUCTION

In recent years many experiments have been undertaken which involve the learning of visual form discriminations by animals: this experimentation, however, has tended to neglect one of the most important aspects of visual form discrimination. By using situations involving such discriminations recent experimenters have for the most part attempted to solve problems of three different kinds: (1) Problems related to the speed of learning when the stimuli to be discriminated are arranged in different ways in relation to the response which will reveal the discrimination, e.g. the successive *v.* simultaneous discrimination issue (Bitterman & Wodinski, 1953), or the approach-avoidance issue (Nissen, 1950). (2) Problems arising from the continuity non-continuity dispute (Blum & Blum, 1949). (3) Neurophysiological issues relating to the site of the centres involved in discrimination in the brain, the specific pathways involved, etc. (Sperry, Miner & Myers, 1955). There is a further problem which in recent years has largely been ignored: this is the problem of arriving at the actual mechanism by which the discrimination is performed. Logically, this question is independent of any of the other three types of question. It is concerned with the way in which visual information is sorted. Questions of type (1) are concerned with the ease with which a given sorting mechanism can be built in or become effective with variations in the ways in which the response must be attached to the stimuli to be discriminated. Experiments of type (2) attempt to answer questions such as, given that an animal has learned to respond in terms of information sorted in one way, how far does this impede or assist the animal in switching to a different method of sorting. Experiments of type (3) have led to some knowledge of the whereabouts of the centres concerned with sorting in the brain: they have given us little information so far about what the sorting mechanism itself is like. One reason for this may be that the functional organization of nerve cells in the brain does not correspond in any simple fashion to their spatial organization: thus comparatively macroscopic lesions may tend to destroy not simply all cells having a similar function in the total organization but cells having a great diversity of function. Thus none of these approaches is specifically designed to throw light on the sorting mechanism *per se*: indeed, for answering questions about the sorting mechanism itself it may not matter whether the sorting mechanism only comes into being through learning or whether it is largely innate—we might be able to answer the question what kind of sorting mechanism was operating in a given animal at a given time without knowing how far it was innate, how far it was learned.

A first requisite for arriving at some idea of what the actual sorting mechanisms used are, is to know what are the properties of the stimuli being discriminated to which the animal is responding when it makes the discrimination. It is also important to know which stimuli the animal finds easy to discriminate and which are difficult for it to discriminate. The explanation of why certain properties are seized upon and not others, and why certain discriminations are easy while others are difficult or impossible, must lie in the peculiar properties of the sorting mechanism. If we know sufficient facts of this kind, we may be able to construct a hypothetical sorting mechanism or mechanisms which will explain the facts. Unfortunately, at the moment we know extremely few of these facts, and nearly all that we do know are due to the pioneer work of a few investigators in the 1930's, and in particular to the classic studies of Lashley (1938) and Fields (1932). There has been little systematic work of this kind in recent years, and the work reported in this paper is intended as a beginning in an attempt to follow up the data assembled by Lashley and others in the thirties, in the hope that by accumulating such data it may eventually be possible to put forward concrete suggestions as to the sorting mechanisms operating in visual discriminations.

In so far as any one factor emerges clearly from the investigations of these problems with the rat in the jumping box conducted in the thirties it is that in this situation one of the properties of figures most readily responded to is that of orientation. Thus Lashley found one of the easiest discriminations for the rat to learn was between vertical and horizontal striations; a rat can distinguish between a triangle on its base and a triangle on its apex more readily than between a triangle and a square; nearly all pattern discriminations are upset by rotations of the figures. If we examine Lashley's evidence carefully, however, a second feature seems to emerge in connexion with discrimination of orientation—this is that discrimination of figures on the basis of their orientation is particularly easy where the main lines of the figures to be discriminated run respectively in the vertical and horizontal directions (vertical versus horizontal striations). Where, however, the two figures to be discriminated are at 90° to one another but do not fall in the vertical or horizontal plane, there is evidence that discrimination may be difficult or impossible. Thus no rat learned to discriminate N from И or S from Ƨ: now the parts of these pairs which differentiate one member of each pair from the other are the lines ╲ *v.* ╱ and ʅ *v.* ʃ. In each case these are identical lines moved round through 90°, but falling midway between the vertical and horizontal directions. Although this evidence is suggestive it is not conclusive: Lashley himself explicitly states that the rat finds no difficulty in discriminating mirror images as such, though Deutsch (1955), apparently relying on Lashley, claims that mirror images are particularly difficult to discriminate. In fact, Lashley's rats had little difficulty in discriminating other pairs of mirror images (e.g. upright versus inverted triangles, upright versus inverted U shapes, etc.).

The present series of experiments is an attempt to test for one lower animal the hypothesis that shapes tend to be classified in terms of their orientation, and that in this classification the extent to which they can be directly related to true vertical and horizontal coordinates is important. There are several reasons why the rat is not a very suitable animal for this study: (1) Lashley found that his rats have a preference for horizontal stripes *v.* vertical stripes, and indeed of his rats which were being trained positive on the horizontal stripes, over half of them learned the discrimination without a single error. Though Lashley himself does not speculate, this may be because in making a jump a rat would normally be jumping to a horizontal surface. Whatever the reason, it complicates the interpretation of these results. (2) Moreover, the rat in the jumping box has been found to tend to respond in terms of parts of the figure and particularly in terms of the lower parts of the figure only (e.g. Munn, 1930), and this again makes results difficult to interpret. (3) In the jumping stand it is difficult to prevent the main lines in the field running in a vertical or horizontal direction: it is seldom clear whether it is the figures themselves which are being discriminated or the figures in relation to the surrounding field (Lashley, 1938).

For these reasons it was decided to conduct the present investigation on octopuses, and also because it was hoped to connect up the writer's own results with the data accumulated on form discrimination and brain lesions in octopus by Young and Boycott. The purpose of this series of experiments then is threefold: (1) To investigate the way in which octopus discriminates orientation, which seems to be one of the most important properties of shapes for discrimination by lower animals, with a view to uncovering something of the mechanism behind discrimination. (2) To compare shape discrimination in the octopus with shape discrimination in the rat. (3) To investigate certain other parameters affecting the behaviour of octopus in the situation used; many of these parameters are also being currently investigated by Young and Boycott, and without a knowledge of their effects it would be impossible to evaluate the data on shape discrimination in the octopus fully. The investigation itself will be reported according to the following plan: (1) An account of the experimental situation and experimental methods used by the present experimenter, including a short account of the habits of the octopus in so far as these are relevant to an understanding of this series of experiments. Fuller accounts of the habits of the octopus will be found in Boycott (1954). (2) An account of the separate experiments conducted and a summary of the results. (3) Discussion of the results.

I. THE EXPERIMENTAL SITUATION

Octopuses were obtained fresh from the sea: they weighed about $\frac{1}{2}$ kg. at the start of the experiments. They were kept individually in separate tanks made of Eternit (a pressed asbestos substance, light grey in colour). The internal diameters of the tanks were about 100 cm. long by 60 cm. wide by 53 cm. high. They were filled with water to a depth of about 37 cm. On the inside the tanks were rounded at all the junctions of sides. At one end of the tank, a 'home' was provided for the octopus made of three bricks against the middle of one short wall of the tank. The octopus spends most of its time sitting in the hole formed by the bricks, looking out down the length of the tank. The tanks were covered by an Eternit lid, which had a hole 25 × 10 cm. cut in it at the end away from the octopus's home to admit light.

Pretraining

The octopuses were given 4 or 5 days in the tank before training started, to allow them to settle down in their new environment.

They were then trained to feed: the lid of the tank was pushed back to a distance of about 25 cm. from the end of the tank opposite the octopus's home, and a crab was thrown in. Though many animals attack the crab readily from the start, particularly if they have been left for several days in the tank without food, others take 10 or 20 sec. to attack at first. Animals were fed five crabs a day in this way at 2 hr. intervals until they reached a criterion of four out of five attacks on 1 day within 5 sec. (Further information on this question will be found in Boycott & Young, 1956*b*). Animals normally reached the criterion of attacking within 2 or 3 days from commencing feeding. Animals which failed to do so were rejected.

When each member of a given group of animals had reached the criterion, the octopuses were then presented with a crab on the end of a holder. The holder was an aluminium tube 7 mm. in diameter and 28 cm. long. This method of presenting the crab was used in preference to the method of Young and Boycott (that of attaching the crab on the end of a piece of cotton and dangling it), because it enables the experimenter to put the crab in contact with the octopus more quickly after the octopus has attacked the positive shape. In fact, it seems doubtful if the slight delay in reinforcement with the method of presenting the crab on the end of a thread makes any difference to the octopus's performance. All octopuses readily attacked the crab on the end of the holder, and only one day's trials were given of this (five trials in all).

Two days' training was then given with the positive shape only, with crab reward. Two animals which failed to attack the positive shape regularly on the second day were discarded.

Positive trials

At this stage and throughout all subsequent stages of the experiments (unless otherwise stated) the method of presenting the positive shape was as follows. Lids were always pushed back about 25 cm. from the end of the tank a few minutes before the experiments began, to give the octopus a chance to adapt to the stronger light. The shapes were cut out of opaque Perspex $\frac{1}{4}$ in. sheet. Since Boycott & Young report difficulty in getting octopus to attack stationary shapes, a moving shape was used. To control the amount of movement, a motor was used to move the shape and the shape was suspended from the motor by a vertical brass rod. The motor imparted an eliptical motion to the shape; its longest movement in any direction was about 2 cm. The experimenter inserted the shape by grasping the motor, and lowering it on to the edge of the tank where it was held firm by a clamp. When in position the shape was 7·5 cm. in front of the end of the tank away from the octopus and the middle of the shape was 14 cm. above the bottom of the tank. The shape was not inserted unless the octopus was in his home at the time or immediately in front of his home. This meant that the shape was presented in front of the octopus, at a distance of about 70–85 cm. from its eye, and slightly above eye level.

The criterion of attack was that the octopus should seize the shape sufficiently firmly to stop the motor revolving within 30 sec. from the entry of the shape into the water. As soon as an octopus had seized the positive shape in this way, it was rewarded with a crab presented directly to the underside of its arms on the end of a holder. If an octopus did not attack within 30 sec., the crab was inserted on the end of the holder and waved about slightly in the neighbourhood of the shape. By keeping the crab slightly behind the shape it was possible to ensure that the octopus did not catch the crab without attacking the shape as well. Normally an octopus which had not attacked the positive shape would attack immediately on being presented with the crab.

Negative trials

In negative trials the shape was introduced in the same manner. If the animal attacked within 30 sec. a shock was administered to the mantle with a shocking device devised by Boycott & Young (1956*b*). This consisted of two electrodes about 1 cm. apart on the end of a long probe. Alternating current at between 6 and 10 V. was used. The electrodes were left in contact with the mantle until the animal released the shape—this usually happened instantaneously. The negative shape was left in for 30 sec. even if it had already been attacked. If the animal made more than one attack within 30 sec. it was again shocked: in fact more than one attack was extremely rare except within the first few presentations of a negative shape.

Training

After pretraining on successively (1) crabs alone, (2) crabs attached to holder, (3) positive shapes, as described above, animals were trained on a discrimination and subsequently tested for various kinds of transference. The procedure on any day which involved training (or retraining) was as follows; five pairs of trials were given a day, at approximately 2 hr. intervals. Each pair of trials consisted of one presentation of the positive shape and one of the negative shape: about 15 min. were left between the two trials making up a pair of trials. On any one day three pairs of trials would have the negative member first, the positive second. Five different orders for pairs of trials were used so that the order of the pairs in which positive came first or in which negative came first varied from day to day: thus the order for the first day (with commas to denote each pair) was + −, − +, − +, + −, − +. In the orders of pairs used no pair of trials with the positive member first was followed by a second pair with the positive member first, and not more than two pairs with the negative member first were placed together. This did not hold good of the order from one day to the next. At the end of each 5 days of training or retraining the order of trials went back to that on the first day. Trials were given in pairs in this way partly for the experimenter's own convenience, partly to make the results comparable with those of Boycott & Young. The effects of this method of presentation and the time intervals used will be discussed later.

Tests

A second procedure was also used at stages throughout the experiments—this we shall refer to as presenting tests. On a day where tests were being given pairs of shapes were presented as above but no reward or punishment was given. On test days the shapes were simply presented to the octopus and removed at 30 sec. or, if an octopus attacked earlier, removed when attacked. We shall refer to the pairs of test shapes used as being positive or negative, though since they were usually shapes not seen by the animal before, and since the animal was not differentially trained on them during test days, they were often not differentially reacted to by the octopus. The order of presentation for all test days was + −, − +, + −, − +, + − for half of any group of animals and − +, + −, − +, + −, − + for the other half. The half of the group getting one order on one test day would get the opposite order on the next test day. On test days, each octopus was

They were then trained to feed: the lid of the tank was pushed back to a distance of about 25 cm. from the end of the tank opposite the octopus's home, and a crab was thrown in. Though many animals attack the crab readily from the start, particularly if they have been left for several days in the tank without food, others take 10 or 20 sec. to attack at first. Animals were fed five crabs a day in this way at 2 hr. intervals until they reached a criterion of four out of five attacks on 1 day within 5 sec. (Further information on this question will be found in Boycott & Young, 1956*b*). Animals normally reached the criterion of attacking within 2 or 3 days from commencing feeding. Animals which failed to do so were rejected.

When each member of a given group of animals had reached the criterion, the octopuses were then presented with a crab on the end of a holder. The holder was an aluminium tube 7 mm. in diameter and 28 cm. long. This method of presenting the crab was used in preference to the method of Young and Boycott (that of attaching the crab on the end of a piece of cotton and dangling it), because it enables the experimenter to put the crab in contact with the octopus more quickly after the octopus has attacked the positive shape. In fact, it seems doubtful if the slight delay in reinforcement with the method of presenting the crab on the end of a thread makes any difference to the octopus's performance. All octopuses readily attacked the crab on the end of the holder, and only one day's trials were given of this (five trials in all).

Two days' training was then given with the positive shape only, with crab reward. Two animals which failed to attack the positive shape regularly on the second day were discarded.

Positive trials

At this stage and throughout all subsequent stages of the experiments (unless otherwise stated) the method of presenting the positive shape was as follows. Lids were always pushed back about 25 cm. from the end of the tank a few minutes before the experiments began, to give the octopus a chance to adapt to the stronger light. The shapes were cut out of opaque Perspex $\frac{1}{4}$ in. sheet. Since Boycott & Young report difficulty in getting octopus to attack stationary shapes, a moving shape was used. To control the amount of movement, a motor was used to move the shape and the shape was suspended from the motor by a vertical brass rod. The motor imparted an eliptical motion to the shape; its longest movement in any direction was about 2 cm. The experimenter inserted the shape by grasping the motor, and lowering it on to the edge of the tank where it was held firm by a clamp. When in position the shape was 7·5 cm. in front of the end of the tank away from the octopus and the middle of the shape was 14 cm. above the bottom of the tank. The shape was not inserted unless the octopus was in his home at the time or immediately in front of his home. This meant that the shape was presented in front of the octopus, at a distance of about 70–85 cm. from its eye, and slightly above eye level.

The criterion of attack was that the octopus should seize the shape sufficiently firmly to stop the motor revolving within 30 sec. from the entry of the shape into the water. As soon as an octopus had seized the positive shape in this way, it was rewarded with a crab presented directly to the underside of its arms on the end of a holder. If an octopus did not attack within 30 sec., the crab was inserted on the end of the holder and waved about slightly in the neighbourhood of the shape. By keeping the crab slightly behind the shape it was possible to ensure that the octopus did not catch the crab without attacking the shape as well. Normally an octopus which had not attacked the positive shape would attack immediately on being presented with the crab.

Negative trials

In negative trials the shape was introduced in the same manner. If the animal attacked within 30 sec. a shock was administered to the mantle with a shocking device devised by Boycott & Young (1956*b*). This consisted of two electrodes about 1 cm. apart on the end of a long probe. Alternating current at between 6 and 10 V. was used. The electrodes were left in contact with the mantle until the animal released the shape—this usually happened instantaneously. The negative shape was left in for 30 sec. even if it had already been attacked. If the animal made more than one attack within 30 sec. it was again shocked: in fact more than one attack was extremely rare except within the first few presentations of a negative shape.

Training

After pretraining on successively (1) crabs alone, (2) crabs attached to holder, (3) positive shapes, as described above, animals were trained on a discrimination and subsequently tested for various kinds of transference. The procedure on any day which involved training (or retraining) was as follows; five pairs of trials were given a day, at approximately 2 hr. intervals. Each pair of trials consisted of one presentation of the positive shape and one of the negative shape: about 15 min. were left between the two trials making up a pair of trials. On any one day three pairs of trials would have the negative member first, the positive second. Five different orders for pairs of trials were used so that the order of the pairs in which positive came first or in which negative came first varied from day to day: thus the order for the first day (with commas to denote each pair) was + −, − +, − +, + −, − +. In the orders of pairs used no pair of trials with the positive member first was followed by a second pair with the positive member first, and not more than two pairs with the negative member first were placed together. This did not hold good of the order from one day to the next. At the end of each 5 days of training or retraining the order of trials went back to that on the first day. Trials were given in pairs in this way partly for the experimenter's own convenience, partly to make the results comparable with those of Boycott & Young. The effects of this method of presentation and the time intervals used will be discussed later.

Tests

A second procedure was also used at stages throughout the experiments—this we shall refer to as presenting tests. On a day where tests were being given pairs of shapes were presented as above but no reward or punishment was given. On test days the shapes were simply presented to the octopus and removed at 30 sec. or, if an octopus attacked earlier, removed when attacked. We shall refer to the pairs of test shapes used as being positive or negative, though since they were usually shapes not seen by the animal before, and since the animal was not differentially trained on them during test days, they were often not differentially reacted to by the octopus. The order of presentation for all test days was + −, − +, + −, − +, + − for half of any group of animals and − +, + −, − +, + −, − + for the other half. The half of the group getting one order on one test day would get the opposite order on the next test day. On test days, each octopus was fed one crab a quarter of an hour after the second presentation of each pair of presentations.

On certain other days during the experiments the animals were allowed to rest: the animals were fed their quota of five crabs on those days.

Throughout the experiments an effort was made to give trials and tests at approximately the same times each day for any animal: for the first animal the times were 10.30, 12.30, 2.30, 4.30, 6.30. Trials were never given at less than 1 hr. 50 min. from one another, but sometimes more than 2 hr. was allowed to elapse between trials in which case the later trials of the day would be correspondingly delayed.

The experiments were conducted during the months of July, August and September and during this time the water in the tanks remained at a temperature of 22° C. ± 1. The crabs used weighed between 8 and 16 g.

In the presentation of the experiments and the results which follow it will be seen that the number of animals used occasionally varies from day to day. This is because during a period of 3 months over which the experiments extend, animals occasionally became unwell. There were two chief causes of ill-health on the part of the octopuses: (1) Octopus climbing out of the tank (often despite the fact that the lid was well secured) and falling on the laboratory floor. (2) Failure of the jets of running sea water in a given tank. In both cases the behavioural effects are a great reduction in the tendency to attack: the scores of octopuses who were unhealthy for one or other of these reasons are not included in the following report for the period over which they remained unhealthy.

Significance of results

The significance of results was tested for in two ways. To find whether any two shapes were differentially reacted to in a series of trials, a χ^2 test was applied to see whether the ratio of number of attacks on the positive shape: failures to attack positive differed from the ratio of attacks on the negative: failures to attack the negative. To find whether discrimination in any given series of tests (A) was significantly better or worse than in any other given series (B), a χ^2 test was applied to the ratio errors made in series A: successes in series A, compared with the same ratio for series B. In all cases Yates's correction for small numbers was applied.

II. THE EXPERIMENTS

There were two major experimental groups of animals. In stage I, group A was trained first to discriminate between a vertical and a horizontal white rectangle 10 × 2 cm.: animals in this group then received training in discriminating rectangles of the same shape but running at 45° to the vertical and horizontal and at 90° to one another. Group B received the same training, but in the reverse order. In stage II various transfer tests based on the discrimination between vertical and horizontal were carried out, with retraining on the original horizontal versus vertical discrimination interspersed amongst the tests. Finally an attempt was made to train the animals to maintain the original discrimination when stationary shapes were used and the rod was removed from the situation. Tables 1 and 2 summarize respectively stages I and II of the experiment. In presenting the results the order in which training and tests were given will not be followed. There were six animals in group A and four animals in group B.

Table 1. *First stage*

Day	Training	Results: No. of trials	Results: Attacks on pos.	Results: Attacks on neg.
	Group A			
1–5	Pretraining	—	—	—
6–12	Training on 10 × 2 rectangles, horizontal *v.* vertical	210	183	54
14–15	Pretraining on positive oblique	—	—	—
16–35	Training on 10 × 2 rectangles obliques	600	332	238
39	Memory test 10 × 2 rectangles, horizontal *v.* vertical	30	25	5
40–42	Retrain, horizontal *v.* vertical	80	68	10
43–44	Training, horizontal *v.* vertical. Stationary shapes	60	46	11
45	Retrain, horizontal *v.* vertical	30	25	6
	Group B			
11–15	Pretraining	—	—	—
16–35	Training on 10 × 2 rectangles obliques	400	194	155
38–39	Pretraining on horizontal *v.* vertical. Positive only	—	—	—
40–45	Training on 10 × 2 rectangles, horizontal *v.* vertical	90	81	24

(1) *Training on the vertical against horizontal discrimination*

This discrimination was between a 10 × 2 cm. white vertical rectangle and a similar horizontal rectangle. It was mastered very rapidly by both group A, which learned it first, and by group B, which learned it after 20 days' training on the oblique discrimination. Fig. 1 gives a graph of the results. The figures for group B are based on three animals only, since one animal from this group escaped from the tank half way through training. Half of the animals in each group were trained positive to the vertical rectangle, half to the horizontal rectangle. There is no evidence that difficulty in learning varied with which rectangle was made positive or negative. Thus, in group A over 7 days, the group trained with horizontal positive made 87 attacks on the positive figure and 27 on the negative, whereas the animals trained with vertical positive made 96 attacks on the positive figure

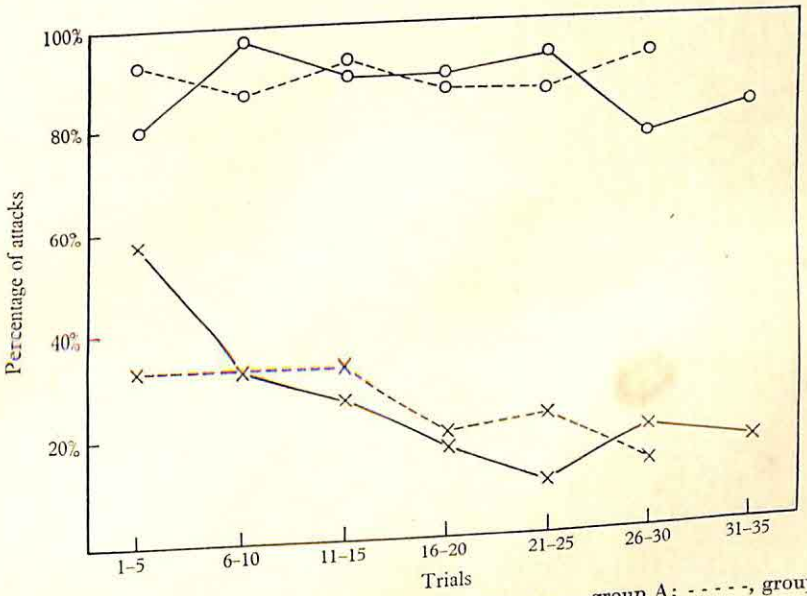


Fig. 1. Learning of vertical *v.* horizontal discrimination. ——, group A; - - - -, group B; ○, attacks on positive; ×, attacks on negative.

and 27 attacks on the negative. It will be noticed that neither group A nor group B became perfect in the discriminations—mistakes were always likely to recur. Although group A received retraining on this discrimination on 15 subsequent days, and group B on 11 days, there was no sign of any general improvement in performance—performance throughout remaining around the level which it had reached on the third or fourth day of training.

(2) *The oblique discrimination*

Training on the oblique discrimination was carried out with the same size rectangles as training on vertical versus horizontal. Half the animals of each group were trained positive on the rectangle sloping one way, half were trained positive on the other. Neither group A (trained on the oblique discrimination after mastering the vertical horizontal discrimination) nor group B (trained on the oblique discrimination first) mastered the discrimination fully. Results are presented graphically in Fig. 2.

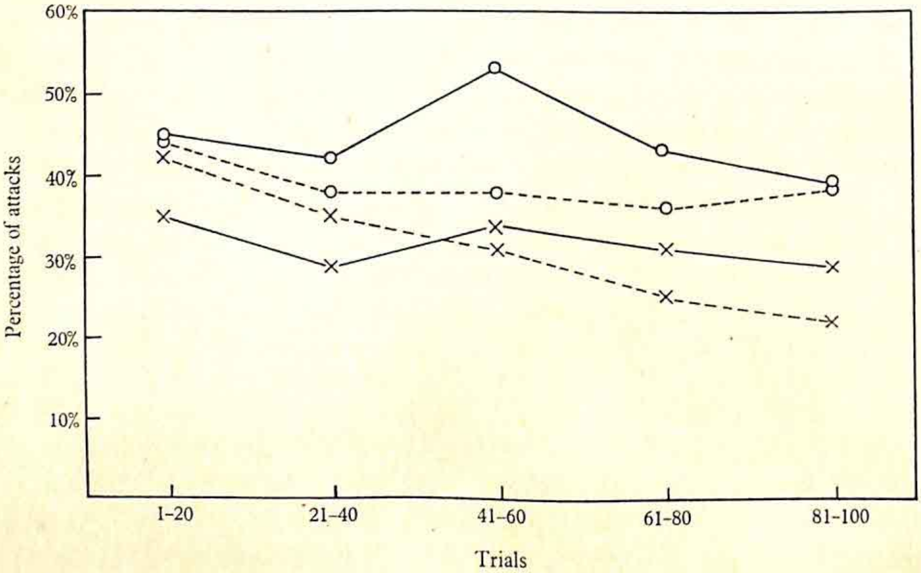


Fig. 2. Learning of oblique discrimination. ——, group A; - - - -, group B; ○, attacks on positive; ×, attacks on negative.

The difference between the learning of the horizontal versus vertical discrimination and the learning of the oblique discrimination is highly significant. For example, when either group A's or group B's performance on the first 2 days' learning of the vertical versus horizontal is compared with either group A's or group B's performance on the first 2 days' learning of the oblique discrimination, the four χ^2 values obtained range from 7·1 to 11·3 with one degree of freedom. Thus the original hypothesis that discrimination of identical figures in different orientations is facilitated when these orientations fall in the vertical and horizontal planes, and is difficult when orientation falls midway between the two is confirmed.

It will be noticed that both for groups A and B the discrimination learned second seems to be learned faster than the same discrimination learned first. Actually the differences on the vertical-horizontal habit are not significant (χ^2 for the first day of A versus the first day of B learning the vertical versus horizontal habit is only 1·7, and for the first 2 days combined is 2·3). The differences on the oblique discrimination become significant only for the first 12 days of learning (χ^2 3·9). What evidence there is suggests that learning of one discrimination slightly facilitates learning of the other.

Individual differences in performance on the oblique discrimination were very considerable; three animals gave no evidence of learning at all, and two animals which had achieved beyond chance scores during the first 10 days make almost an equal number of attacks on the positive and negative figures during the second 10 days. This falling off in discrimination in a difficult habit has been noticed by Wells & Wells (1956) with tactile discrimination in the octopus.

(3) *Further experiments with original vertical versus horizontal figures*

(*a*) *Movement.* It may be suggested that although movement was controlled by a machine, the original vertical versus horizontal discrimination may have been based on slight differences in the motions of the figures due to their differential resistance to water pressure when moving in different directions. To control this possibility group A were retrained on days 43–44 with stationary shapes: the shapes were presented as before, but the motor was not started. The scores produced on these two days (46 attacks on the positive, 11 attacks on the negative out of 60 presentations) do not differ significantly from scores on days when the shapes were in motion.

(*b*) *Presence of the vertical rod.* It is possible that the vertical versus horizontal discrimination was being performed by the animals by reference to the vertical rod, at the end of which the shape was attached. However, controls used rendered this unlikely. Thus substitution of a transparent Perspex rod for the original brass rod did not affect the discrimination. In addition, an attempt was made to see whether the animals would continue to discriminate the vertical from the horizontal when the rod was no longer present in the situation. The positive and negative shapes were presented one at a time, attached to a board of the same size as the far wall of the tank against which the shapes were normally seen and made of a grey Eternit substance similar to the walls of the tank. This board was lowered vertically into the tank, just inside the wall of the tank away from the octopus's home, and when it came to rest at the bottom the shape was in the same position in the tank as it had been when presented on the end of a rod. One animal continued to discriminate (8 attacks on the positive, 1 on the negative during the first 10 pairs of trials), but the other animals showed reduced tendency to attack, or attacked the board rather than the shape. It seems likely that the size of the board may have been producing fear in some of the animals (colour changes associated with fear were frequently observed), and also the shape was no longer treated as an independent object but as part of the board with which of course it moved as the board was lowered into the tank. Despite this some discrimination was continued: the score on the first day was 13 attacks on the positive and 3 on the negative out of 45 trials with each (9 animals), on the last 2 days together 34 attacks on the positive and 12 on the negative out of 60 trials with each (6 animals). This indicates at least that the vertical rod is not an essential cue in the discrimination of vertical versus horizontal.

(4) *Memory*

Boycott & Young (1955, 1956*a*) suggest that if the shapes learned are not shown, some forgetting of a discrimination may occur comparatively rapidly in octopus. One test of retention over a period of time was included in the present series of experiments. On day 39 group A was tested for retention of the vertical horizontal-discrimination, after a period of 27 days without being called on to make this discrimination. The horizontal and vertical rectangles were those that the animals had been trained on. The score for 6 animals was 25 attacks on the positive and 5 on the negative out of a total of 30 presentations in each case. Although no reward or punishment was given, this in fact is the same score as on the last day of training with these shapes. It may be, of course, that intervening training on the oblique discrimination had helped to preserve the memory for the original horizontal-vertical discrimination. Since, however, the oblique rectangles were both exactly midway between the horizontal and vertical positions, this is unlikely.

(5) *Transfer tests*

The procedure employed during transfer tests has already been described. Each transfer test was always preceded by at least one day's retraining on the original discrimination. Since in the original training half the animals were trained positive to the vertical, half to the horizontal shape, in transfer tests where positive results were obtained the positive of the two transfer shapes for any animal was determined by the shape to which it was originally trained positive. Transfer was tested with seven different pairs of shapes. To guard against the possibility that the animals were becoming 'experimentally wise' on days when transfer tests were given, the same shapes were used on two separate days—once towards the beginning of the transfer tests and once towards the end. The similarity of the scores obtained for the two days indicates that other differences amongst results in transfer tests cannot be explained in this way (see Table 2, transfer test 2 and transfer test 7).

Table 2. *Second stage, transfer tests*

Figures used	No. of trials	Attacks on pos.	Attacks on neg.	χ^2 for significance of discrimination	χ^2 for significance of difference from normal discrimination
1. Black 10 × 2 rectangles, vertical *v.* horizontal	45	36	32	N.S.	16·9*
2. 7 × 2 rectangles, vertical *v.* horizontal	45	36	3	*	N.S.
3. 20 × 4 rectangles, vertical *v.* horizontal	45	36	4	*	N.S.
4. 5 × 1 rectangles, vertical *v.* horizontal	40	23	11	6·2*	13·6*
5. 4 × 2 rectangles, vertical *v.* horizontal	40	22	13	3·3	12·9*
6. Equilateral triangles, side 10 cm. One side horizontal *v.* one side vertical	45	21	17	N.S.	16·2*
7. 7 × 2 rectangles, vertical *v.* horizontal	40	31	5	*	N.S.
8. 7 × 1 rectangles, vertical *v.* horizontal	40	28	8	18·2*	N.S.

A χ^2 value of 3·8 is significant at the 0·05 level of confidence.
N.S., not significant.
* Significant.

N.B. (1) Each test day was preceded by at least 1 day's retraining. (2) The first column of χ^2 values indicates the extent to which the difference in number of attacks on negative and positive shapes for any given test day is significant. (3) The second column of χ^2 values indicates whether there is a significant difference between number of errors made on any given test day and errors made on the preceding day's retraining. (4) Where χ^2 values are less than 2·0 or greater than 20·0, exact values are not shown.

The results of the transfer tests are summarized in Table 2. Two statistics were computed on the scores for each transfer test: the first column of χ^2 values gives the result of an analysis conducted to see whether the two shapes were differently treated by the animals in any given test. The second column gives the significance of the difference between number of errors made in a transfer test and number of errors made on the day of retraining preceding each test. Scores on days of retraining did not vary significantly from one to another—if the retraining day when most errors were made (18 out of 90 trials) is compared with the day when least errors were made (9 out of 90 trials), the significance of χ^2 is only 2·8.

The conclusions to be drawn from the transfer tests may be summarized as follows (numbers in parentheses indicate the number of the transfer test referred to in Table 2):

Where white rectangles of different sizes from the originals were substituted, some transfer was found in all cases. Transfer was very much better for rectangles four times the area of the originals than for rectangles one-quarter the area of the originals (3, 4). Perfect transfer occurred where the ratio of long to short sides was altered from 10:2 to 7:2 or 7:1 (2, 7, 8): although some discrimination remained, transfer was significantly less where the ratio was altered to 4:2 (5), though in this case some of the reduction in transfer may be due to the lesser area of the figures (cf. 4). It seems unlikely that the relatively poor transfer to smaller figures can be explained in terms of visual acuity, since where there was little or no discrimination (as in the case of the oblique figures and the small figures under discussion) the octopus would launch its attack from a position much closer to the figure than where there was good discrimination: also octopuses seem to have no difficulty in identifying pieces of fish much smaller than the smallest figures used here.

From the results of the transference tests to figures of different sizes, it can be concluded that the octopuses were not discriminating on the basis of the relationship between the rectangles and the immediate surround. Thus they could not have been reacting on the basis of the distance between the lower edge of the rectangle and the bottom of the tank, since the lower edge of the *horizontal* 20 × 4 rectangle is the same distance from the bottom of the tank as the lower edge of the *vertical* 4 × 2 rectangle and yet they react differentially to these forms (3, 5). A similar argument shows that the octopus is not reacting in terms of the absolute lateral extent of the rectangles—the lateral extent of the *horizontal* 4 × 2 rectangles is the same as the lateral extent of the *vertical* 20 × 4 rectangle. Thus these transfer tests establish clearly that the octopus is reacting to the ratio between long and short sides of the rectangles and is not discriminating in terms of any one absolute distance involved in the situation.

The test with equilateral triangles (6) indicates that the octopus is not discriminating merely on the orientation of one long side, since there is no transfer from vertical and horizontal rectangles to two triangles, one with a side in the horizontal plane, the other with a side in the vertical plane.

As is the case with rats, there is little or no transfer when black forms are substituted for the original white ones (1).

(6) *Simultaneous discrimination*

After the transfer tests had been given, animals were retrained to discriminate between the original vertical and horizontal rectangles without the presence of the rod. This training is described in (3*b*) of the present section. It was thought worth while to follow this training to respond differentially to rectangles presented one at a time on the board, with training in which both rectangles were presented simultaneously on the board. Animals were shocked if they touched the negative shape, and fed a crab if they touched the positive shape. Six animals were used at this stage of the experiment and training lasted for 2 days. (Shapes were presented to the left and right of the board at the same height that shapes had been presented when given singly.) This stage of the experiment lasted 2 days, making a total of 50 trials. On 26 trials animals attacked only the positive shape, on 9 trials they attacked either the negative shape or both shapes together. Three animals did not touch the negative shape throughout. Thus it seems that octopuses, having learned differential reactions to two shapes presented separately, may be able to transfer the reactions to the same shapes presented simultaneously.

(7) *Sequential effects*

It might be thought that in the original horizontal versus vertical, and oblique discriminations, the octopuses may have been learning to attack or not attack the second shape of a pair on the basis of what had happened with the first shape of the pair. It will be remembered that the presentations of the positive and negative members of a pair were separated by only a 15 min. interval. An octopus could in theory learn that if it had had a crab in connexion with the first member of a pair, then it must avoid the second member of a pair; if it had not had a crab in association with the first member of a pair then it was safe to attack the second member of a pair. That this was not the basis of the discrimination is proved for the vertical versus horizontal discrimination, by the fact that in the memory tests and in certain transfer tests animals produce scores as high as during retraining despite the fact that no reward or punishment was given in connexion with each member of any pair. To prove that the discrimination which appeared in the oblique discrimination was not based simply on making reactions to the second member of a pair conditional on whether or not the first member had been rewarded with a crab, it is necessary to analyse the results and show that the different number of attacks on negative and positive is not based wholly on a difference between numbers of attacks on the second members of a pair. These results are presented in Table 3. It will be seen that discrimination is as good for the first members of a pair as for the second members of a pair though total number of attacks are reduced for both negative and positive shapes when these were the second members of a pair. A full discussion of the analysis given in Table 3 will follow in the Discussion.

Table 3. *Breakdown of scores on oblique discrimination*

(Figures represent percentage of attacks.)

	%		%
Total positive	53	Total negative	39
First positive	61	First negative	43
First positive before attack on negative	81	First negative before attack on positive	48
First positive before no attack on negative	50	First negative before no attack on positive	38
Second positive	47	Second negative	34
Second positive after attack on negative	53	Second negative after attack on positive	46
Second positive after no attack on negative	43	Second negative after no attack on positive	15

III. DISCUSSION

Discrimination of orientation

The results show that the octopus can discriminate much more readily between a vertical and a horizontal rectangle than between two opposite oblique rectangles at 45° to the vertical and horizontal. The difference in ease of discrimination does not depend upon differences of movement nor upon the presence of the vertical rod in the situation. Transfer tests have eliminated the possibility that the octopus is discriminating on the basis of the immediate relation between the forms and their background, and have shown that the octopus is discriminating the different ratios between horizontal and vertical extent produced by the vertical and horizontal rectangles.

Although the immediate background of the forms has been excluded as providing a frame of reference in terms of which the octopus might make the discrimination, the possibility that it is making the discrimination with reference to some wider property of the background has not been excluded: the vertical axis is clearly marked in the situation used by a number of asymmetries between top and bottom of the tank (direction of light, presence of floor of tank at bottom, water surface at top, etc.). It is possible to specify the difference between the vertical and horizontal rectangle in terms of the vertical co-ordinate alone—it is not possible to specify the difference between the two obliques in terms solely of this co-ordinate: thus if we wish to measure the angle of tilt of the two oblique rectangles from the vertical plane, in order to distinguish between them we must specify whether the angle is to be measured in a clockwise or anticlockwise direction and this involves reference to a second co-ordinate, e.g. to left and right on the horizontal co-ordinate. It is of course also possible that the vertical axis is marked by information from statocysts in a way in which the horizontal axis is not marked. Since it is extremely difficult (and perhaps impossible) to remove all indicators of the vertical axis from the situation, experiments to discover whether the difference in ease of discrimination between the two pairs of rectangles would disappear if the vertical axis were not clearly demarcated have not been performed. In fact, whether the vertical and/or horizontal axes are of importance because of the arrangement of the external environment, or because of the arrangement of the octopus's nervous system may not make much difference to the type of system that underlies the classification of forms: if the environmental difference is important, it must itself be reflected somewhere in the octopus's nervous system.

It is now proposed to fulfil the promise made earlier in this paper and show how behavioural evidence may enable us to put forward hypotheses concerning the system responsible for classifying shapes. Two alternative hypotheses are put forward to explain the present results: neither is developed in detail since at the moment it is impossible to choose between them, and the writer hopes to develop them in detail in a subsequent paper on the basis of further evidence. A crucial test for distinguishing between the two hypotheses is advanced.

(1) One hypothesis is that the octopus's nervous system analyses the lateral extent of forms, while preserving the position of the parts of the form on the vertical axis. This would mean that points on the vertical axis had a specific representation in the nervous system: this specific representation, if not innate, is likely to have developed as a result of learning since throughout the animal's life there are present clear and consistent differences in the environment (direction of light, etc.) corresponding to differences of height: this is not the case for differences along the horizontal axis. When a horizontal rectangle is analysed in this way, strong excitation (corresponding to its lateral length) would be produced along a small band of the points in the nervous system representing differences along the vertical axis: when a vertical rectangle is analysed in this way weak excitation (corresponding to its small lateral extent) would be produced over a large band of the points representing differences along the vertical axis. The two oblique rectangles would

produce identical excitation patterns since both their lateral and vertical extents are the same. This system would account for the difference in ease of discrimination found, and is easy to translate into neural terms. In order to perform the observed transference to rectangles of different sizes, all that is necessary is that instead of analysing the absolute amount of excitation produced by the lateral extent of the figure, this amount is analysed relative to the total area of the figure.

(2) The other possibility is that the octopus's nervous system analyses both lateral and vertical extent of the figures. In other words that both lateral and vertical positions in the visual field have specific representations. Such a system would again discriminate readily between vertical and horizontal rectangles (vertical—strong excitation over narrow band of points on lateral axis, weak excitation over broad band of points on vertical axis; horizontal—the reverse of this), and again it would give the same output for both oblique rectangles (medium excitation over a medium band of points both on the lateral and vertical axes). The second system would perform the size generalizations more easily than the first, since all that would be required is that the amount of excitation over any given range of points on the vertical and lateral axes be compared: if it is stronger on the lateral axis the rectangle is vertical, if on the vertical axis the rectangle is horizontal.

A crucial test for the two systems is as follows: the second system should discriminate equally well between a pair of T figures orientated ⊥ and T and a pair orientated ⊣ and ⊢: the first system should discriminate much more readily between the former pair than the latter. Lashley, in fact, found rats discriminated more readily between U figures differing in the vertical axis (upright and inverted) than between U figures differing in the lateral axis (figures lying on opposite sides): but because of the rat's habit of discriminating in terms of the base-line of the figure, this finding is difficult to interpret.

The fact that octopuses can with much training discriminate slightly above chance between the oblique rectangles does not contra-indicate these hypotheses: it would be possible to perform this discrimination on the systems proposed if the octopus first learned to fixate a given part of the figure. One of the advantages of both systems is that they will explain recognition of many orientations independent of point of fixation. These hypotheses, and particularly the second, are capable of explaining many facts of form vision: they will be developed in more detail, if further experimental testing seems to warrant it.

Fluctuations in tendency to attack

The reduction in the number of attacks made both for positive and negative shapes when the shapes were presented as the second members of a pair as compared with numbers of attacks made on the first members of a pair requires some explanation. The figures on which the following analysis is based are presented in Table 3. They are figures for all animals learning the oblique discrimination. Similar trends appear, however, in the training, retraining and testing with the vertical and horizontal shapes, but the trends there are not so pronounced since they are overshadowed by the effects due to good discrimination of the vertical and horizontal figures.

The situation is complicated because tendency to attack seems to have varied from one pair of trials to another in each octopus. We can see this by comparing the percentages of attacks on the positive when the positive came first for animals who went on to attack the negative against the percentage for animals who did not go on to attack the negative. The percentages are 81 versus 50—a highly significant difference. This is not a serial effect since the two groups are selected on the basis of how they reacted to the *second* figure of the pair. The difference in these percentages represents fluctuation in some general tendency to attack. There is a similar difference in percentages of attacks on the first negative between animals who went on to attack the positive and those that did not (48 *v.* 38%). We would of course expect such fluctuations in general tendency to attack merely on the basis of generalization from one figure to the other apart from other considerations.

However, this fluctuation in general level of tendency to attack will not explain the reduced number of attacks on the second members of pairs of figures. The obvious explanation for reduction in number of attacks where the positive figure was second is to suppose that this is due to the reduction in attacks in octopuses who had just attacked the negative and had received a shock. To find out whether this is the case or not we must compare two sets of figures: (1) (*a*). Percentage of attacks on second positive where the negative had been attacked, with (*b*) percentage of attacks on first positive where the following negative was attacked. (2) (*a*). Percentage of attacks on second positive where the negative had not been attacked, with (*b*) percentage of attacks on first positive where the following negative was not attacked. We must make the comparison in this way in order to parcel out any effect due to fluctuations in general tendency to attack: thus 2 (*a*) cannot be compared directly with percentage of animals attacking the first positive, because 2 (*a*) is already selected as a group which does not attack the negative and this in itself might tend to produce a lower percentage of attacks in the second positive. Now when the comparisons are made in this way we get the following figures: (1*a*) 53%, (1*b*) 81%; (2*a*) 43%, (2*b*) 50%. The difference in comparison (1) is highly significant, that on (2) falls just short of significance ($\chi^2 = 3{\cdot}1$). The fact that the first difference is greater than the second suggests that the shock is reducing the number of attacks on the second positive. The fact that there exists a nearly significant difference in (2) suggests that some additional factor may have to be taken into account in explaining the difference in number of attacks on the second positive from number of attacks on the first positive. This impression is strengthened when we break down the results for percentage of attacks on the second negative.

Making these comparisons for percentage of attacks on second negative figures in the same way as for the second positive to avoid effects due to general fluctuations in level of attack, we have: (3*a*) percentage attacks on second negative after an attack on positive, compared with (*b*) percentage attacks on first negative before an attack on positive. (4*a*) percentage attacks on second negative after no attack on positive, (*b*) percentage attacks on first negative before an attack on positive. The percentages are (3*a*) 46, (3*b*) 48, (4*a*) 15, (4*b*) 38. Once again percentages of attacks on the second figure are down as compared with the first, though this time they are only significantly down where the positive figure was not attacked within 30 sec. It must be remembered that even in cases where the positive figure was not attacked within 30 sec. it was attacked after 30 sec. when the crab was shown in association with it. It may have been decreases in motivation due to getting a crab that led to the difference in (4), but if this is so a similar difference should show in (3). The main difference between animals who did attack the positive first (3*a*) and animals who did not attack the positive first (4*a*) will be that the positive shape has been exposed longer for ($4a$) than for ($3a$). All our results can be explained if exposure to a shape tends to reduce the tendency to attack a shape—in comparison (1) there will be the additional factor of the effects of shock.

Now this possibility is of considerable importance in the context of Boycott & Young's experiments on the effects of lesions to the verticalis (Boycott & Young, 1955). They have found a tendency for animals lacking the verticalis to lose discrimination and to be unable to relearn it under normal conditions. They tend to attack both positive and negative shapes. However, some of the animals can be retrained by presenting shapes at very short intervals—a condition under which normal animals either do not learn at all or are at least much retarded since they cease attacking. In an experiment not yet published, Wells & Wells present similar findings for touch discrimination in animals lacking the verticalis. Young & Boycott suggest tentatively that the negative memory in verticalis animals is very short lived and can only be kept going if the traces are strengthened by frequent presentation (every 5 or 10 min.) of the negative shape even when the negative is not attacked. The present results suggest as a further possibility that verticalis removal may raise general tendency to attack and that discrimination can only show up when this tendency is reduced, and that this reduction may be brought about by very frequent presentation of shapes.

Reality of situation

The rates of learning of the octopus in the situation used seem to compare very favourably with rates of discrimination learning for the rat. It may be that this is because the situation used (as developed by Boycott & Young) is much closer to what the animal is called on to do in its natural environment than are discrimination situations hitherto used with rats. In this connexion the following features should be noted: (1) The reaction is made directly to the shape to be discriminated: Murphy & Miller (1955) found considerable reduction in rate of learning for monkeys where the response was not made directly to the object discriminated. (2) The reaction is a natural reaction for the octopus and does not involve elaborate motor learning. (3) The reward is one that the octopus would normally get for making a positive reaction to shapes in its natural environment. (4) The shapes are the shapes of objects, not merely two-dimensional patterns. It seems likely that animals are normally called on to discriminate shapes of objects not patterns and that object discrimination may be an easier discrimination for animals to make than discrimination between two-dimensional patterns. Moreover, it may be that because the shapes are real objects (emphasized by their movement against the rest of the field) that the animals seem to transfer something learned about this shape in one context (on the end of a rod) to other contexts (successive and simultaneous discriminations with the shapes on a board).

The writer wishes to thank the Royal Society for making the work possible by a grant from the Browne Research Fund, Prof. J. Z. Young for much advice and encouragement, Dr Brian Boycott for much helpful advice, Mr Martin Wells for advice about and help with both apparatus and animals, and the Director and Staff of the Stazione Zoologica for the facilities and hospitality accorded to him.

REFERENCES

BITTERMAN, M. E. & WODINSKI, J. (1953). Simultaneous and successive discrimination. *Psychol. Rev.* **60**, 371–7.

BLUM, R. A. & BLUM, J. S. (1949). Factual issues in the 'continuity' controversy. *Psychol. Rev.* **56**, 33–50.

BOYCOTT, B. B. (1954). Learning in *Octopus vulgaris* and other Cephalopods. *Publ. Staz. Zool. Napoli*, **25**, 67–93.

BOYCOTT, B. B. & YOUNG, J. Z. (1955). A memory system in *Octopus vulgaris* Lamark. *Proc. Roy. Soc.* B, **143**, 449–80.

BOYCOTT, B. B. & YOUNG, J. Z. (1956*a*). Reactions to shape in *Octopus vulgaris*. *Proc. Zool. Soc. Lond.* **126**, 491–547.

BOYCOTT, B. B. & YOUNG, J. Z. (1956*b*). Memories controlling attacks on food objects by *Octopus vulgaris* Lamark. *Pubbl. Staz. Zool. Napoli*, **27**, 232–49.

DEUTSCH, J. A. (1955). A theory of shape recognition. *Brit. J. Psychol.* **46**, 30–7.

FIELDS, P. E. (1932). Studies in concept formation. I. The development of the concept of triangularity by the white rat. *Comp. Psychol. Monogr.* **9**, no. 2, 1–70.

LASHLEY, K. S. (1938). The mechanism of vision. XV. Preliminary studies of the rat's capacity for detail vision. *J. Gen. Psychol.* **18**, 123–93.

MUNN, N. L. (1930). A note on Lashley's method for studying vision in the white rat. *J. Gen. Psychol.* **37**, 528–30.

MURPHY, J. V. & MILLER, R. E. (1955). The effect of spatial contiguity of cue and reward in the object quality learning of rhesus monkeys. *J. Comp. Physiol. Psychol.* **48**, 221–4.

NISSEN, H. W. (1950). Description of the learned response in discrimination behaviour. *Psychol. Rev.* **57**, 121–31.

SPERRY, R. W., MINER, N. & MYERS, R. E. (1955). Visual pattern perception following subpial slicing and tantalum wire implantations in the visual cortex. *J. Comp. Physiol. Psychol.* **48**, 50–8.

WELLS, M. J. & WELLS, J. (1956). Tactile discrimination and the behaviour of the blind octopus. *Publ. Staz. Zool. Napoli*, **28**, 94–126.

(*Manuscript received* 6 *December* 1955)

CORRIGENDUM

In 'Some studies of perception in relation to age' by Jean G. Wallace, Vol. 47, Part 4, page 287, Table 2, under 'Group 2', should read as follows:

Age	No. successes	No. failures				No. failing
		1	2	3	4	
51–60	18	2	6	.	.	8
61–70	17	1	1	.	3	5

This correction affects neither the correlation nor the Figure on page 287 since both are derived from the correct data.

PUBLICATIONS RECENTLY RECEIVED

The Organization of the Cerebral Cortex. By D. A. SHOLL. London: Methuen. 1956. Pp. xvi+125. 18*s*.

This admirable little book by the Reader in Anatomy at University College, London, should find a place in every psychological library. Indeed, I would make it obligatory reading for every student of psychology proposing to embark on the fashionable grand tour through the Hull and Hebb country.

After a brief account of the available methods of histological study, Dr Sholl proceeds to outline the earlier work on the structure of the cortex. The outcome is devastating to anyone brought up in the tradition of Brodmann, Campbell and the Vogts. On cortical stratification, for instance, Dr Sholl writes: 'It is clear that each observer has his own qualitative and subjective criteria for his schema of lamination and there is no agreement as to the definition to be adopted for a discrete cortical lamina' (p. 19). Cytoarchitectonic maps fare little better. Although it can be accepted that certain regions of the cortex show differences in histological structure, '...attempts to subdivide these main regions on the basis of criteria subjectively applied to preparations stained by a selective technique are unsuccessful' (p. 28). This conclusion may well cause many of us to regard the standard texts on the physiology of the brain with a warier eye. Indeed, Dr Sholl comments that the conventional criteria of structural differentiation, taken in isolation from any consideration of connectivity, are unlikely to be correlated at all closely with the essential activity of the central nervous system.

Dr Sholl next considers the quantitative approach to the histology of the cortex—an approach largely due to Bok and his followers and to which the author has himself made important contributions. This approach deals with such attributes of structure as the total thickness and extent of the cortex, the total number of neurons and the density of neuronal packing, the number and relative thickness of the laminations, and the proportions of specific types of neurons in the different cortical areas. It is clear that a good many facts about these and related problems are now known at the gross statistical level, but technical difficulties remain very great. From the point of view of an understanding of function, the outcome of quantitative neurohistology is not as yet very impressive.

Following a most useful discussion of neural connectivity, Dr Sholl examines briefly some aspects of cortical organization as revealed by neurophysiological and clinical inquiry. This part of the book is perhaps less satisfactory than the rest, largely owing to limited coverage and drastic compression. It is concluded that whereas parts of the cortex, owing to their two-way connexions with subcortical structures may well be associated with specific types of behaviour, '...the main volume of cortex cannot be described in terms of specific "functions"' (p. 80). One may point out that this conclusion, though doubtless acceptable to Lashley, is perhaps misleading unless appropriately qualified. In so far as clinical material is concerned, it is noteworthy that the author limits himself almost entirely to operations on the frontal lobes, the effects of which are notoriously difficult to evaluate. Had he considered comparable data relating to the parietal or temporal lobes, the impression given might well have been very different. True, nothing is harder than to define a cortical 'function' by inference from the derangements produced by cortical lesions. But unless the effort is made it is doubtful whether cerebral neurology can advance further.

In his later chapters Dr Sholl surveys briefly the mode of operation of the cortex in the light of some current theories. The Pavlov theory is dismissed as totally inadequate. The theories of Lashley and Hebb are regarded more favourably, though certain difficulties of both are examined. Hebb, in particular, is taken to task for his undue reliance on the data furnished by strychnine neuronography and, more generally, for failure to provide a quantitative application of his theory. (No doubt Dr Hebb would be the first to agree with these criticisms.) It is, however, pleasing to note that Hebb's book is considered as '...possibly the only one that presents psychological problems in a way that can be appreciated by the anatomist and physiologist' (p. 91).

Cybernetic theories likewise come under fire. On Ashby: 'It is true that examples of adaptive behaviour in animals are given that do not conflict with the notion of ultrastability, but there is no evidence that, in fact, the central nervous system does embody this principle and no indication is given as to what anatomical and physiological facts correspond with the design' (p. 93). On Grey Walter: 'It is most unlikely that the unique complicated circuits suggested by the author's diagrams exist in any nervous system' (p. 94). Biologists, Dr Sholl urges, are interested not only in what an animal does but in how in fact it does it. Psychologists attracted by electronic models could do worse than ponder his words.

Dr Sholl's positive contributions to theory are modest but far from trivial. In view of his interest in the quantitative and statistical approach to cerebral structure, he not unnaturally leans towards a quantitative and statistical treatment of cerebral function. He is much impressed by recent developments in mathematics and their application to biological problems. He sees hope, for instance, in the sort of ideas recently put forward by Beurle and by Uttley, though is modest enough to state that a theory recently advanced by Uttley and himself '...seems to demand a cortical organization for which there is no adequate evidence' (p. 105). Although not everyone will agree that the theory of games offers the best basis for an understanding of behaviour, it is to be most earnestly hoped that Dr Sholl and his associates will carry their ideas as far as they will possibly go.

This little book is 'the genuine article'. It is expert and informed, restrained yet adventurous. It should prove an astringent corrective to much nonsense talked about brain function by those who should know better. Psychologists—who don't always know better—should read the book with particular care. What if Pavlov does turn in his grave? Dr Sholl is only putting into practice Pavlov's own dictum that in the study of the nervous system every step of thought should be verified through agreement with actual fact. And if neurons are not facts, what are?

O. L. ZANGWILL

Learning and Instinct in Animals. By W. H. THORPE. London: Methuen. 1956. Pp. viii + 493. 55*s*.

Dr Thorpe is exceptionally well qualified for writing a non-partisan text on animal behaviour. His extensive knowledge of ethological, psychological and naturalistic source materials enables him to put together and often to integrate a great diversity of facts and theories. The result is a book that will reorient any reader and provide him with a wealth of new information.

The book has three parts. The first contains an examination of the concepts of 'directiveness' and 'purposiveness', and a critique of instinct theory. The point of view adopted is purposive and organizational. Cognitive processes are emphasized and it becomes clear that the author finds it congenial to use a terminology of expectancies, hypotheses and the like. A general drive to explore and perceive the environment is postulated. Perceptions may be goals. Drives may be endogenous in the sense that they arise with the C.N.S. itself. The review of instinct theory justifies the acceptance, with some reservations, of Lorenz's notion of action specific energy. Tinbergen's hierarchical schema of energy channels is criticized because of its rigidity and, to account for the 'flexible directiveness of appetitive behaviour', Dr Thorpe finds it necessary to postulate some process of working towards 'ideal' or 'full' releasers.

In Part II some of the features of habituation, conditioning, trial and error learning, latent learning and insightful learning are examined. Imprinting is not given the status of a distinct form of learning. The discussion of habituation is particularly valuable since this topic has not received from psychologists the attention it deserves. Latent learning is found to involve a simultaneous reorganization and integration of past experiences and thus to some extent to be insightful. The experiments of the rat laboratory gain in significance by being discussed in the general context of orientation behaviour as it occurs in a variety of animal forms, invertebrate as well as vertebrate. A final chapter deals with some recent speculations on the neurophysiological mechanisms of learning.

Part III, more than half of the book, is headed 'The learning abilities of the main animal groups', but the range is wider than this may indicate. Birds and arthropods receive the fullest treatment in chapters that are extensions and revisions of previously published journal articles. The new material includes some interesting studies of the solution of string-pulling problems by birds. It is perhaps a little disappointing to find only 16 pages alloted to protozoa, coelenterates and echinoderms. The chapter on mammals is also short, but justifiably so, and the principle is adopted that mammalian behaviour should be regarded as learned unless proved otherwise. Throughout this part of the book experiments have been selected for mention because of their relevance to the biology and natural history of the animals concerned. This ensures that learning is not treated as an isolated process. The classification adopted earlier in the book proves to be convenient, whatever other justification it may have.

Dr Thorpe's final comments concern the unity of behaviour as it appears in diverse animal forms and the unity of the physical and biological sciences.

Much of the book consists of short descriptions of experimental work coupled with shrewd interpretative comments. Theories are dealt with in critique rather than in expository form. The level of difficulty makes the book unsuitable for the beginner but excellent as an advanced text or reference. In one place or another, attitudes are expressed towards all the main problems of behaviour. While a consistent perceptual-organizational point of view is maintained, there is no attempt at formal theory construction. On occasion the absence of Occam's razor is noticeable, but not to be regretted in this case since a more constrictive approach would have reduced the scope of the book and deprived the reader of some valuable insights.

The merits of the book as a whole are so outstanding that criticism tends to be inhibited. However, there are three sources for dissatisfaction. First, terms like 'instinct', 'purpose' and 'drive' are not used consistently. Secondly, the reader too frequently meets assertions that cannot be interpreted with any precision even in their context, for example, '...perception is a basic characteristic of the drive of the living animal'. Thirdly, the point of certain theories seems to have been missed and in places unconsidered criticisms are made. For example (p. 103), we are told that the only possible interpretation of avoidance training is that an 'anticipation' (author's quotes) of the noxious stimulus has been acquired, and that to postulate 'fear' as a drive is a superficial stratagem because the term connotes anticipation (without quotes) by definition unless the James-Lange theory is accepted.

The publication of this book is an important and timely event for all students of animal behaviour. It is to be hoped that it finds the place it deserves on the bookshelves of psychologists and ethologists alike, and that its integrative approach and its provision of common background information will help to promote a more unified outlook.

R. L REID

Analysis of Perception. By J. R. SMYTHIES. London: Routledge and Kegan Paul. 1956. Pp. xiii + 140. 21*s.*

The most remarkable feature of this very interesting book is that it combines two approaches, the neurological and the philosophical, which are usually kept separate. The author is a neurologist who has made a careful and detailed study of recent philosophical literature. As a philosopher he is able to clarify the important but obscure concept of 'the body-image': and as a neurologist he is able to point out the weaknesses of current philosophical theories of perception. He is also an expert on hallucinations, especially those induced by mescaline, and his chapter on 'Veridical and Hallucinatory Sense-experiences' is one of the best in the book. His main aim, however, is to show that 'a synthesis of the sense-datum theory with the neurological account of perception may constitute, if its basic terms can receive adequate definition, the only valid theory of perception in both science and philosophy' (p. 4).

He defines the term 'sense-datum' itself by reference to after-sensations: roughly, a sense-datum is an entity such that it or some part of it can have a common boundary with an after-sensation. (The term 'after-sensation', he says, needs no formal definition; it can be defined operationally and ostensively.) Sense-data are spatial entities; they can even be measured, e.g. by reference to the total diameter of the visual field. In these respects, there is no difference between hallucinatory sense-data and normal ones. Here, and throughout, Dr Smythies is mainly concerned with *visual* sense-data, though he does have something to say about somatic sense-data as well. His version of the sense-datum theory could no doubt be applied to the other sense-modalities, but some additional definitions would obviously be needed. He then argues that the Naive Realist theory, which tries to dispense with sense-data altogether, is inconsistent with the empirical facts about the finite velocity of light. The theory of 'psycho-neural identity', which identifies sense-data with physiological events in the brain, is rejected on the ground that the relevant cortical events are not geometrically congruent with the sense-data alleged to be identical with them, while the obscure doctrine of the 'projection of sensations' only makes confusion worse confounded. Dr Smythies therefore accepts a dualistic interaction theory. But it differs from the traditional Cartesian theory in two important respects. According to Dr Smythies, the mind is itself an *extended* entity, because the extended sense-data which it senses are actually constituents of it. On the other hand, physical space is not the only space there is. There are also a number of sensible spaces, a different one for each percipient, in which sense-data are located. Thus we have a choice between two unorthodox cosmological theories: in the first, there are no spatial relations between sense-fields and physical objects, and then we have a very complicated universe in which there are many different spaces; in the second there is an all-embracing n-dimensional space, of which physical space and the various sensible spaces are three-dimensional sections (where n, if I understand rightly, $=6$). Dr Smythies, I think, is inclined to prefer the n-dimensional theory; it would follow that psycho-physical interaction must occur across a 'dimensional interface'. He also suggests, however, that at some future time we may be able to devise experiments which will make it possible to decide between the two theories on empirical grounds.

Dr Smythies frankly calls his theory of perception a representative theory. The great difficulty of representative theories, from Locke's time until now, is that they seem to make it impossible to us to know anything about the physical world. Dr Smythies tries to meet this by drawing two closely connected distinctions: (1) between sensing and perceiving, (2) between direct and indirect observation. He holds that we do really perceive the physical objects we suppose ourselves to perceive: the sensing of a sense-datum is only a part of the total process of perceiving, not the whole of it. Again, sense-data are indeed the only entities we observe *directly*; but in so doing, we *indirectly* observe the physical objects which they represent.

But if sensing is only one part of the total process of perceiving, what is the other part? The other part consists in the operation of a very complex 'non-symbolic representative mechanism'. This ensures that the spatial properties of sense-data will correspond pretty closely with the spatial properties of the physical objects which set the mechanism going by stimulating the sense-organs. Dr Smythies claims that we can get light on the nature of this mechanism by studying stroboscopic phenomena, which suggest strongly that it works on much the same principles as a television apparatus (ch. 2). Because our nervous system does contain a pretty reliable representative mechanism of this kind, we are justified in maintaining that by direct observation of sense-data, we are indirectly observing physical objects.

However interesting and suggestive this may be from a causal point of view, I cannot think it gets over the epistemological difficulty. Dr Smythies, like Locke before him, does not seem to have distinguished sufficiently clearly between two different situations: (1) x senses a sense-datum s which is *in fact* produced by a representative mechanism affected by the object 0, and the spatial properties of x do in fact correspond with the spatial properties of 0. This might be called the *de facto* sense of the word 'perceive'. (2) When x senses a sense-datum s produced in this way, he is thereby enabled to *discover* or *find out*, non-inferentially, that 0 exists and has these spatial properties. This might be called the cognitive sense of the word 'perceive'. However true the propositions listed under (1) may be, we need to establish the truth of (2) as well, if the classical difficulties of the representative theory of perception are to be solved.

H. H. PRICE

Fundamental Statistics in Psychology and Education. By J. P. GUILFORD. New York and London: McGraw-Hill. Third edition, 1956. Pp. xi+565. 47*s*.

The Measurement of Abilities. By P. E. VERNON. London: University of London Press Ltd. Second edition, 1956. Pp. xii+276. 20*s*.

Although very different in scope these two new editions may conveniently be noticed together, in view of similarities both in subject-matter and in the principles underlying the revision. Each is quite clearly the same work as before, but brought up to date by the inclusion of a small proportion of fresh material. In each case, too, the text has been entirely reset, with more printed matter to the page, and consequently a smaller total number of pages.

The difference in scope probably explains why, at first sight at least, there appear to be greater changes in Guilford's book after six years than in Vernon's after sixteen. Recent refinements in statistical method and changes in emphasis have allowed Guilford to discard some things that appeared in earlier editions: thus Probable Error—which Vernon continues to treat fairly fully—is dismissed in a very brief footnote and not even defined. The disappearance of certain other material, notably the chapter on scaling procedures, is explained by its having been transferred to the author's *Psychometric Methods* (review in this *Journal*, 1955, **46**, 145–6). (Similar transfers having taken place in the reverse direction, successive editions of the two works would seem to be progressing—out of phase, so to speak—along parallel paths.) Another relatively major change is the expansion of the previous single chapter on Testing Hypotheses into two. Guilford comments on the progress that has been made in this area, but maintains his bias towards descriptive rather than sampling statistics, on the grounds that 'research in the social sciences is not confined to studies in which results are at the margin of significance', and that 'the generation of scientific ideas...does not depend particularly upon decisions concerning change alternatives'. There are several useful new tables, including one for reading off cosine-pi correlation coefficients from values of *ad/bc*.

Modifications in the statistical chapters in Vernon's book are, necessarily, less extensive, but it is worth noting that such topics as chi-squared (and degrees of freedom generally) and small-sample statistics are dealt with more adequately than was possible at the time the book first appeared. It is in the sphere of the theory and practice of intelligence testing, particularly in relation to current problems in educational administration and research, that the more important changes have been made. The work thus retains its unique character as a comprehensive survey of all aspects of testing and measurement in education, including traditional examinations, and its genuine comprehensibility to the non-mathematician has not been attained at the cost of misleading over-simplification. The invaluable classified list of tests available in this country, now completely reconstructed, is exemplary for the clarity of its layout.

BORIS SEMEONOFF

Discussions on Child Development. Vol. I. edited by J. M. TANNER and B. INHELDER. London: Tavistock. 1956. Pp. 240. 25*s*.

This praiseworthy volume augurs well for the series of which it is the first. In 1953 the World Health Organization brought together a number of distinguished scholars to discuss, in the words of the book's subtitle, 'the biological, psychological, and cultural approaches to the understanding of human development and behaviour'. This study group contained seventeen members and represented the fields of psychology, psychiatry, ethology, cultural anthropology, electro-physiology, and psychoanalysis: it included, to name but a few, John Bowlby, Bärbel Inhelder, Konrad Lorenz, Margaret Mead, Jean Piaget and Grey Walter. This book gives a running account of the group's discussions and the editor must be congratulated on having 'tidied up' the transcript without depriving it of its friendliness and informality. The structuring of the discussion itself represents a worthwhile experiment in inter-disciplinary communication. To begin with, each participant introduces himself by giving a brief autobiography and a sketch of his special interests. Then follow eight sessions, each of which is opened by a paper from one member on some aspect of development. The ensuing discussions are the important thing and, as might be expected from the calibre of the participants, they are erudite, wide-ranging, and stimulating: in them, the same or analogous phenomena are looked at from varying viewpoints, and there emerge a host of interesting facts and exciting ideas. It would be impossible to select any one portion of the discussions for particular mention: suffice it to say that there was hardly a page which did not provide the reviewer with some unthought of parallel or some new idea. The aim of the meeting was not to read papers, pass resolutions, or issue reports but to discuss human behaviour in a permissive and multi-disciplinary atmosphere and achieve, thereby, some measure of understanding and cross-fertilization between the several disciplines which converge to study behaviour. For once, a meeting of this kind can fairly be said to have achieved its aim and the book's many readers will be grateful to assume the role of eaves-dropper. The reviewer can think of no psychologist, however specialized his interests, who will not derive both pleasure and profit from this volume.

I. M. L. HUNTER

Alfred Adler: an Introduction to his Psychology. By LEWIS WAY. Harmondsworth: Penguin. 1956. Pp. 252. 3*s*. 6*d*.

This latest addition to the Pelican Psychology Series contains a brief biographical sketch of Adler, and an outline of his theories, illustrated by many extracts from his writings. It will be interesting and useful both to the student and the general reader. Though he was a first-rate speaker, Adler was a diffuse and repetitive writer, and he is an excellent subject for the boiling-down process to which he is here subjected.

The book should dispose finally of certain misconceptions about Adler, as that he was a Nietzschean power-worshipper who regarded agression as the dominant human motive. 'The striving of every living individual', he wrote, 'is directed towards achievement, not towards power'. *Gemeinschaftsgefühl* or 'social interest' is one of the basic Adlerian concepts; and the neurotic, in Adler's view, is essentially a person who is unable to achieve satisfactory relations with others owing to the self-absorption and touchiness produced by inferiority-feeling.

Adler called his system *Individualpsychologie* to express his view that the person must be considered as an individual whole—that symptoms cannot be treated in isolation, but must always be seen in relation to the total personality and to the style of life. He was not an individual psychologist in the current sense; in fact, he was a pioneer of social psychology. In his first published work, written when he was 28, he said: 'The doctor can no longer ignore the standpoint which examines men, not as isolated units, but as products of their society.' Later, he was one of the first to point out the importance of role and culture-pattern in the formation of personality. Like most psychologists of his day, he was apt to make sweeping generalizations, prefaced by some such phrase as 'I have observed' or 'it is evident'. But he had a remarkable knack of hitting the nail on the head, and many of his pronouncements have been amply confirmed by later investigation. To take only one example, his statements about the effects of deprivation of affection in early childhood (cf. pp. 101–2) provide a striking anticipation of the work of Bowlby.

MARGARET KNIGHT

Introduction to Psychology. By CLIFFORD T. MORGAN. New York and London: McGraw-Hill. 1956. Pp. vii+676. 45*s*.

The choice of an introductory text for students of general psychology is influenced by many considerations, the least mundane being the chooser's theory of education. There is much to be said for each of the possible approaches—the historical, the biological, the methodological, the broad survey of the whole field, and so on. This book is of special interest since Prof. Morgan has recently devoted his time to introductory teaching. His conclusion seems to be that a text should 'give them the facts'. In this remarkably comprehensive survey of the contemporary scene controversial issues are avowedly omitted. There are four pages on historical background, including the 'schools', and six on methodology (as opposed to techniques), but in the rest of the book there is little or no attempt to relate findings critically to past theory or present method. Although occasionally the reader is told that one experiment is good or better than another, he has to take the author's word for it.

The book has been arranged so that only selected chapters need be read depending on which of three courses the student is following a 'life oriented' course, a 'science oriented' course, or a comprehensive course. In fact the whole book is life oriented in that findings are continually related to everyday experience. Various chapters have been drafted by fourteen psychologists, but the main author has rewritten them all to avoid stylistic changes. The result is extremely readable, and, as far as this reviewer knows, accurate. The book would be appreciated by the readers of a good popular science periodical. The sections on physiology are somewhat more learned, which might be expected under this author. The lists of further reading at the end of each chapter are really excellent. About one quarter of the space is taken up by pictures and tables, many of them gorgeous, and many unnecessary. There is a glossary which contains 'apes' and 'brain' but not, for instance 'ambivert', 'catharsis' 'cerebellum' and 'central tendency', which are all used in the book.

Those psychologists who take psychology to be a suitable subject for a liberal education, in that it provides exercise in clear thinking and stimulates intellectual excitement, may not be enamoured of this text. However, those who wish to prescribe a simple, clear, up-to-date and thoroughly comprehensive summary of present-day psychology will find it at least the equal of several other popular introductory texts, and more recent than most. The chapters comprising the 'science oriented course' might prove particularly suitable for those reading psychology as part of a general degree.

B. M. FOSS

Rank Correlation Methods. By M. G. KENDALL. London: Griffin. Second edition, 1955. Pp. vii+196. 40*s*.

This is a welcome re-issue, with much new material, of Prof. Kendall's important work, first published in 1948 and apparently little known to the majority of psychologists, although its influence on design and treatment in recent published experiments has been considerable.

Although the scope of the book is comprehensive, the main content is an exposition of the theory and use of the rank correlation coefficient τ (tau), which has the advantage over ρ that its significance can be estimated with much greater accuracy. It is, however, more cumbersome to calculate, since it involves comparing ranks for all possible pairs of objects or individuals. This number of operations, i.e. $\frac{1}{2}n(n-1)$, quickly rises to formidable proportions for any but a very small sample, and indeed certain of the appendix tables, as well as the demonstrations in the text, seem to imply that samples of 10, or at the very most 20, will be the largest requiring to be handled. Further, obtained values of τ are always less than those of ρ; it is consequently of less value as an estimate of product-moment correlation.

An additional merit of τ, however, is its generality, i.e. it can be extended to a wide variety of purposes, such as biserial and partial correlation and paired comparisons. While recognizing that 'a tool which has many purposes is not usually as efficient for any one of them as a specialized tool developed solely for a single purpose', the author sums up by pointing out that 'there are situations where the power of rank-order tests is surprisingly high and little seems to be lost by employing ranking methods'. There are undoubtedly circumstances, particularly among those involving evaluative judgments, in which Prof. Kendall's methods are the most appropriate of those available, and no statistically-minded psychologist can afford to neglect to add them to his repertoire. Since most of even the new and recently revised statistical texts are curiously reticent on the subject, the present book is indispensable.

BORIS SEMEONOFF

Opinions and Personality. By M. BREWSTER SMITH, JEROME S. BRUNER and ROBERT W. WHITE. London: Chapman and Hall. 1956. Pp. vii+294. 48*s*.

In 1947 the authors of this book, three American Social psychologists, induced ten men from different walks of life—a law clerk, a salesman, a factory operative, a real estate agent, and so on—to submit for two hours a week for fifteen weeks to a battery of tests and interviews. The object was to find out as much as possible, from their twenty-eight 'procedures', about the personalities of their subjects, and from this to judge the way in which their opinions on a particular topic fit into the personality pattern which the investigators had constructed for each of them. The topic chosen was 'Russia', a topic on which everyone may be expected to hold some opinion, and one about which many people feel strongly. All the subjects except one were hostile in varying degrees; the exception, while not being a member of the Communist Party, was strongly in favour of the Soviet regime. In passing it may be said that it was unfortunate that the investigators did not manage to find more people in this position.

The project was intended to be something of a corrective to the simpler statistical approach. In the view of the writers, the correlation of a single trait or constellation of traits with opinions 'pro' or 'con' is all too simple. The object towards which the attitude is felt is itself complex. In this case 'the Russia to which each referred was a conception selectively fashioned, a reflexion of individuality'. Moreover, a hostile or favourable attitude may function in one way for one person and in another way for another.

There are, it is suggested, three main functions served by opinions and attitudes. First, 'each person is constantly trying to size up the world around him and to place it in relation to his major interests'. Attitudes, in fashionable jargon, are 'structuring' agencies. Secondly, a man is influenced by such need as he may have to conform, or deliberately not to conform. Thirdly, attitudes may be vehicles for the expression of inner problems. The upshot is that an opinion may be considered as 'a resultant or compromise between reality demands, social demands and inner psychological demands'. Furthermore, they are influenced by the availability of information, by intellectual ability, and by the holder's 'expressive style'—whether, for instance, he is a man of theory or a man of action.

All this is worked out in detail for three of the subjects, and in less detail for the other seven. The case for multiple causality is, as we should expect, made out. The case histories, indeed, are models of exposition. The results, as I have suggested, are not surprising, but it is well that we should be reminded of a complexity all too often forgotten. Perhaps the most significant aspect of the results is the implications they carry with respect to attitude change. If, for example, a man holds certain views about Russia simply because he is interested in the subject, then his views may be changed by further information. If, on the other hand, his views serve to make him feel he is in the right company, or if he is expressing some inner tension, then information will be of no use, and other methods must be used. Propagandists must take into account the variety of vested interests which may lend support to any opinion they want to alter.

W. J. H. SPROTT

The Psychology of Industrial Conflict. By ROSS STAGNER. New York: Wiley. London: Chapman and Hall. 1956. Pp. ix+550. 64*s*.

Human Relations in the Industrial Southeast. By GLENN GILMAN. North Carolina University Press. London: Cumberlege. Pp. xii+327. 40*s*.

Textbooks and case-studies of 'human relations in industry' continue to pour from the presses. It is good to see that the latest volumes represent a distinct improvement on the earlier ones—the ritual chant 'workers are human beings!' which formed much of the substance of the early message of the human-relations movement has now given place to a sober examination of institutional demands upon the individual (including the influences of unions and employers' associations), the part played by historical factors in industrial conflict, and the diverse motivational patterns of occupational groups. There has also been a refreshing tendency for industrial research to be brought within the ambit of sociological and psychological theory. Dr Gilman, for example, offers the reader an extended industrial case-study, presented in sociological terms, while Prof. Stagner attempts the first comprehensive and systematic psychological approach to industrial conflict.

As a text-book, *The Psychology of Industrial Conflict* is first-rate. The book is well laid out, so that one is led from one chapter to the next by a systematic line of thought. It opens with a survey of research on perception and motivation, first in their general aspects (with the emphasis on social differences) and then in their application to industrial life. This is followed by chapters on frustration and aggression, social groups, institutions and leadership, in each of which the relevant industrial material is presented within a framework of concepts which relates it (somewhat eclectically, it is true) to research work going on outside the industrial field. Later chapters take up the industrial theme more directly: management and union tactics, the strike, and industrial co-operation. The coverage of published research is remarkably complete, including British as well as American work up to 1954. The result is a rather fat text-book of over five hundred pages, which can be taken, by those who like an approach on a broad front, in conjunction with Kornhauser's two volumes to provide a comprehensive introduction to an important group of problems.

Dr Gilman is a Chicago-trained sociologist specializing in industrial management at the Georgia Institute of Technology. His book is in part a case-study, describing the growth of the textile firms located in south-eastern U.S.A., in the southern Piedmont region. If he had gone no further he would have produced an interesting historical study of an important American industry, limited in appeal to those who are concerned with textiles. But Dr Gilman has chosen to present his material 'against a frame of reference that is sufficiently fundamental to serve the same purpose for many industries in many regions'. The frame of reference is provided by the 'work, place, people' concept of Geddes, derived from Le Play. This is developed into an attempt to interpret the history of the Piedmont textile industry as an expression of basic human nature (compound of 'needs for understanding', security, and novelty), available natural resources and available cultural resources. Inevitably, since this is primarily a sociological study, the psychological treatment is rather rough-hewn. On the cultural level, a useful working distinction is made, between a folk-society and a mass-society (substantially the distinction made by Tönnies between *Gemeinschaft* and *Gesellschaft*). Eighteen differences between the two social types are combined in a check-list, and the Piedmont region is seen to be a folk society, emphasizing solidarity, diffuse rather than specific ties between individuals, and homogeneity rather than heterogeneity. Industry in such a society takes on a characteristic form, differing in many ways from industry of the same type in a mass-society. This long and very detailed case-study forms a useful addition to research on the 'community factor in modern technology'. At the end, the author has included a couple of dozen photographs of modern textile plants and their surroundings, showing the physical appearance of mill communities.

J. F. MORRIS

A Follow-up Study of War Neuroses. By N. Q. BRILL and G. W. BEEBE. Washington: Veterans Administration. 1955. Pp. xviii+393. No price given.

The authors report a follow-up study, conducted under the auspices of the Veteran's Administration, of a sample of 1475 white men (excluding officers) of the U.S. armed forces who had missed duty during 1944 with a diagnosis of psychoneurosis. Of these cases, about one-third were studied from records only, while about two-thirds were interviewed by a psychiatrist approximately five years after their breakdown in 1944. Two hundred and twenty-five psychiatrists took part. Information about two control groups was used, one of enlisted men in 1944, the other of enlisted men in 1951. A large number of factors, connected with both service experience and pre-service experience and personality, are considered in relation to breakdown and to mental health at the time of follow-up. The findings are set out in a very large number of tables. Many of the assessments must be regarded as rough and tentative, but in the main the results are more or less what one would expect. An interesting point is that, at the time of follow-up, about 40% of the men felt that they still needed treatment. But 'there is no evidence that treatment played an important role in the general improvement which occurred between separation and follow-up'. In the concluding pages, the authors express the view that 'predisposition has been overgeneralized, and that we would do well to think more in terms of vulnerability to specific forms of stress and to the factors that may modify that vulnerability', that is, attention should be directed to the effective use of those with neurotic tendencies, and not merely to trying to keep them as far as possible out of the forces.

D. GRAHAM

Self-evaluation and Rejection in Groups. By JOACHIM ISRAEL: Uppsala: Almquist and Wiksell. 1956. Pp. 250. Sw. kr. 20.

This monograph consists of a review of the literature, a 'conceptual outline', and reports of three laboratory experiments. It is very much in the Michigan tradition and draws heavily on the experiments and theories of Leon Festinger and his group. The first experiment is the Swedish part of the cross-cultural study of rejection of deviates, directed by Schachter. The results show insignificant tendencies in the predicted direction, though the deviates *were* rejected here. The second experiment tests some of the author's hypotheses concerning rejection of deviates of different ability in co-operative and competitive groups, and these are largely confirmed. The third experiment shows some of the conditions under which the group's evaluation is accepted as opposed to a conflicting self-evaluation. All three experiments were skilfully conducted and carefully interpreted though the write-up is very hard to follow.

The review of the literature and 'conceptual outline' are unusual in the amount of discussion given to questions of definition, not all of which is illuminating. A series of hypotheses is developed, but these are not deductively related as in Festinger's theory on which they are based: each is asserted independently, following a short discussion of the problem.

MICHAEL ARGYLE

Performance Tests of Intelligence under Indian Conditions. By C. M. BHATIA. London: Oxford University Press. 1955. Pp. xi+131. 17*s.* 6*d.*

This short book, with a foreword by Sir Godfrey Thomson, is a straightforward account of the construction and standardization of a battery of individual performance tests suitable for use with Indian children. The battery consists of part of the familiar Kohs Block Designs and Alexander's Passalong Test, together with tests of Pattern Drawing, Immediate Memory and Picture Construction, devised by the author.

The battery has been administered by Dr Bhatia and his co-workers to representative samples of literate and illiterate Indian children ranging from 10 to 16 years of age and a separate conversion table drawn up for each of the two groups of children. There is an interesting, though brief, account of the special measures taken by the testers to establish good relations in the villages visited, a prime necessity in obtaining a suitable atmosphere for testing, much of which seems to have been carried out in the open with an audience. A more detailed account of the difficulties met with in testing might have helped to augment the Western reader's insight into the Indian scene.

Not unexpectedly, the mean performance of literates is found to be superior to that of illiterates on every test. The author maintains that illiteracy in India is related, not to differences of race or of culture, but to lack of educational opportunity. His two independent standardizations for the two groups reflect his insistence that valid comparisons are restricted to subjects within each group separately, and that no such inter-subject comparisons should be made across the groups. The chapter on 'The Nature of Intelligence' is sketchy and might well be expanded.

A. E. G. PILLINER

Allgemeine Theorie der menschlichen Haltung und Bewegung. By F. J. J. BUYTENDIJK. Berlin: Springer. 1956. Pp. 367. D.M. 39·60.

This is a German translation, and therefore a more generally accessible version, of the author's *Algemene Theorie der menselijke Houding en Beweging*, published in the Netherlands in 1948. It considers human posture and movement from a point of view intended to collate physiological and psychological data. Human bodily activities, must, it is claimed, be regarded as 'functions', which differ from 'processes' in being 'indivisible whole(s), consisting of changes which are meaningfully related to something outside themselves'. The something outside may be, for example, a goal to which an activity is directed and which gives it its value. Furthermore, human behaviour must be seen as the behaviour of a *self*, which is neither solely physical nor solely mental but partakes of both realms in a psycho-physical unity. Not only associationism and the Pavlovian type of reflex theory but also Gestalt psychology, with which Buytendijk's position has much in common, are blamed for leaving no room for the 'Subject', which 'disposes of its own body', 'manifests itself as the cause of the perceived movement' but is not localizable in the nervous system.

The questions most contemporary psychologists would like to ask about human posture and movement—in what conditions are particular sorts of movement most likely to be performed? How can movements be predicted and controlled? What is the precise nature of the neuromuscular mechanisms underlying movement?—receive comparatively little attention, apart from assorted data mentioned as evidence for the main theses of the book. There is, in fact, little of what most of the readers of this *Journal* think of as psychology. On the other hand, there is a great deal of phenomenology, the method of inquiry introduced by Husserl and extended by many continental philosophers, notably the existentialists, which leads to various conclusions from the fact that we cannot help thinking about things in certain ways.

It is interesting to ask what impression of psychology the reader of the book will obtain. One fact that will not spring at him is that most of the psychology in the world, for good or for ill, has for some time been the product of English-speaking countries. Nor will he see that recent years have been at least as prolific as any others of innovatory developments. Remarkably few of Buytendijk's references are later than 1939, and almost all of those that are come from continental Europe. The reader will not be given the impression that there are many contemporary theories in psychology which agree with Buytendijk that behaviour does not depend solely on external stimuli of the moment, that it is often purposeful and that various pieces of behaviour and neural processes interact, and yet do not accept Buytendijk's somewhat vitalistic conclusions. Finally, it must be admitted that Buytendijk disdains any meretricious devices that might encourage people to think that psychology is all froth and glitter. The text is written in a somewhat viscous style and is unrelieved by any illustrations, tables or graphs. It is hard to believe that anybody whose staple diet is neither German philosophical literature nor suet pudding will find the book an exhilarating one to read.

D. E. BERLYNE

A New Psychotherapy in Schizophrenia. By M. SECHEHAYE. New York and London: Grune and Stratton. 1956. Pp. vii +199. $4.50 net.

In two previous publications (*Symbolic Realisation* and *The Autobiography of a Schizophrenic Girl*) Mme Sechehaye described her highly original and successful treatment of a young schizophrenic girl by a psychotherapeutic method which was called 'Symbolic Realisation'. The present publication is essentially her attempt to order her previous observations into a systematic theory which will allow others to apply her method in treating schizophrenic patients.

Although the author offers a completely psychological mode of treatment she views schizophrenia as the result of the impact of early trauma on a constitutionally predisposed schizoid individual. She thus accepts the concept of a 'latent schizophrenia' which may become actualized by the frustration of fundamental needs, tolerated by the non-predisposed individual. The method of symbolic realization sets the therapist the task of satisfying these unrelieved needs at a symbolic level appropriate to the genetically primitive ego state of the patient and thus 'deliver him from the awful guilt holding him captive and driving him to his death'. The author's obvious empathy with the schizophrenic patient allows the reader to accept the frequently flamboyant style of expression which might otherwise have been objectionable. The vivid account of her patient's therapeutic progress leaves one in no doubt as to Mme Sechehaye's ability to enter the schizophrenic's 'world of experience' (the existential Erlebniswelt). It is, however, difficult to assess the validity of her case for 'symbolic realization' as a curative agent. Its essentially subjective and often obscure nature suffers from the added difficulty that Mme Sechehaye relies primarily on retrospective reports of the now non-psychotic patient to verify the validity of her early interpretations of the patient's psychotic behaviour. An even more fundamental weakness lies in the fact that the method and underlying theory is based solely on the author's treatment of one schizophrenic patient and, as the author admits, 'it remains to be seen whether utilizing the same method in the same way, under conditions as alike as possible, and with similar symptoms, the same phenomena of cure will appear again'.

However, the value of this book cannot be judged by the validity and applicability of the author's therapeutic method. In an attempt to understand the subjective world of the schizophrenic Mme Sechehaye utilizes concepts from the diverse fields of psychoanalysis, existential phenomenology and the genetic psychology of Piaget. It is impossible to mention all the stimulating ideas raised by the author who has obviously a deep and sympathetic understanding of the schizophrenic patient. Although most of her information is gathered from her experiences with one patient, there is much to be said for the assertion of Goldstein and others that the intensive and detailed study of one patient is often of more significance than a more superficial study of many. Whether or not we are impressed by Mme Sechehaye's therapeutic method, no one who professes an interest in the prime clinical problem of schizophrenia can afford to neglect consideration of the author's absorbing observations.

A. MCGHIE

Les Mécanismes Cérébraux de la Prise de Conscience. By PAUL CHAUCHARD: Paris: Masson. 1956. Pp. 240. 1300 frs.

Dr Chauchard attempts to achieve a biological conception of consciousness based largely on neurophysiological considerations. His masters are Lapicque and Pavlov, but he is conversant with a good deal of the relevant data from clinical neurology and psychiatry. He is antagonistic to both classical dualism and the new materialism of Cybernetics, seeking an understanding of consciousness in terms of the evolution of the central nervous system itself. So far as the present reviewer can judge, nothing particularly original emerges from his study. None the less, it can be commended as an essay of good standing in the French physiological tradition.

O. L. ZANGWILL

Emotional Hazards in Animals and Man. By HOWARD S. LIDDELL. Springfield: Thomas. Oxford: Blackwell. 1956. Pp. xi+97. 20*s.*

Dr Liddell has been studying abnormal behaviour in animals for some thirty years, and this little book, based on a series of lectures given to a psychiatric audience in Halifax, Nova Scotia, represents the outcome of his experience. After a vivid description of a demonstration of conditioned foreleg flexion by a sheep, a chapter on the reassuring effects of the presence of the mother during the conditioning of young animals follows. Liddell then discusses those aspects of the conditioning procedure which he considers the most stressful for animal subjects, and which he terms loneliness, monotony, confusion and overstimulation. The manifestations of 'neurotic' behaviour in his experimental animals are then described, to which some notes by Mrs Prochazka on the effects of faulty training procedures in horses are appended. In the fifth chapter the emotional response of the experimental animal to its own interpretation of conditioning signals is discussed, and stages in the onset of emotional breakdown are identified. In his final chapter, Liddell elaborates his observations of the behaviour of front-line soldiers under battle conditions in the light of the behaviour of his experimentally disturbed animals, and adds some comments on the role of pleasurable emotions in mental health.

Much of the material gathered together in these lectures is familiar from Liddell's previous writings. The style is conversational, and, while experimental findings are frequently cited, they serve to illustrate rather than to convince. Anecdotal evidence is not eschewed. There is no index, but a selected bibliography of nineteen references is provided.

The author's viewpoint has changed over the years, and this change towards his present emphasis on the importance of the conditioning situation as such in the aetiology of emotional disturbance is epitomized in his statement that 'the conditioning routine which we regarded as innocuous apparently meant murder to the sheep'. But his criteria of what constitute an 'experimental neurosis' are loosely defined, and, indeed, appear in this book to fluctuate.

It is difficult, therefore, to see to what audience the book is addressed; experimental psychologists will find it insufficiently organized to serve as a useful treatment of Liddell's undoubtedly major contribution to the experimental literature in this field. This reviewer's bias would lead him to attribute this to Liddell's failure to provide or employ a theoretical system which might illuminate his empirical data and be enhanced in turn thereby. On the other hand, the book does not appear to offer psychiatrists much insight into the part played by emotional strain in the development of neurotic disorders, beyond some simple animal analogues. Some, however, may find these satisfying and helpful.

P. L. BROADHURST

The Art of Making Sense. By LIONEL RUBY. London: Davies. 1956. Pp. 285. 16*s.*

This is one more of those books which aim at improving the reader's mind by the administration of formal logic in graduated doses, heavily diluted with anecdote. Of its kind it is reputably done. Unlike some authors in this field, Prof. Ruby avoids the pretence of identifying logic with the psychology of thinking, nor does he attempt to peddle the subject as a royal road to truth. On certain points he is undeniably superficial, notably in his off-hand treatment of verbal meaning and casually uncritical acceptance of the correspondence theory of truth, but on the whole he covers the ground fairly, within the limits that a beginner can be expected to understand.

The opening chapters are mainly concerned to pin down argumentative misdemeanours and non-formal fallacies of semantic or psychological origin. This is followed by a straightforward account of formal reasoning, syllogistic and hypothetical, supplemented by a fleeting glimpse of symbolic methods and a rather curious defence of the law of excluded middle against the clamours of Korzybski and his followers—unnecessary at this level, one would have thought, and certainly out of place in an English edition, for the non-Aristotelian Count is by no means the celebrity here that he is (or was) in the States. The remainder of the book is devoted to probability, evidence, induction and scientific method, with a final chapter on value judgements; the sections on gambling, legal inference and 'the logic of Sherlock Holmes' (a remarkably ill-informed theorist, by the way) are all interestingly put over, though English readers might well have been given a footnote on the rules of 'craps' and readers of all kinds may wonder why, with the whole history of science at his disposal, the author has chosen to pick his examples of 'how the scientist thinks' from Dr Cronin's *Citadel*.

The first duty of a popular logician is to keep his audience awake. Prof. Ruby knows this, and sparkles away conscientiously at about three corny stories to the page; all the same, one cannot help feeling that his book would be even more readable if it were less obviously anxious not to be dull.

P. L. HEATH

The Nature of Sympathy. By MAX SCHELER. Translated by Peter Heath. London: Routledge and Kegan Paul. 1954. Pp. liv+274. 30s.

This addition to the series 'Rare Masterpieces of Philosophy and Science' is of less direct interest to psychologists than its title would suggest. On the other hand, when time and inclination offer it is rarely unprofitable to explore somewhat alien rumination. Whether Scheler's book is a masterpiece or even the secondary writing of a master seems to the present reviewer open to question. It has, however, historical interest and in the light of Gardner Murphy's recent speculations on the limits of personality it even has some topical interest—its very confusions reveal the pitfalls awaiting the unwary. Quite unquestionable is the value of Dr Stark's excellent introduction.

At the time of writing 'Sympathy', Scheler argued that there are various levels of human understanding and up to what he calls the 'vital' level there is no real problem about our sympathy with or knowledge of other people because direct awareness is primary. This looks interesting until one realizes the pantheistic context. In 1913, still retaining the Thomism he later abandoned, Scheler went on to maintain that there are spiritual as well as vital levels of existence and understanding. And at the spiritual level there is an essence or core of every personality which, if it is ever known, will only be apprehended through a very subtle form of 'fellow feeling'. This for Scheler presupposes an equally subtle notion of loving and recognizing another person as forever a distinct and independent entity.

Scheler tries introspectively to distinguish other varieties of relation apt to be left unscrutinized over here because English, unlike German, has no ready-made words like *mitleid* or *mitfreude*. Difficult though he is, Scheler can be persuasive and his discussion is occasionally illuminated by flashes of insight that give one pause. But one should not be blinded by artistry. And torn from the context of what in any case seems deeply confused philosophizing Scheler's views on direct awareness, for example, hardly stand up to psychological scrutiny (beyond his sensible attempt to defend some element of social responsiveness as primary). Moreover, in the cold clear light of an English morning the idea that other people merit love, fellow feeling or respect as independent entities appears either vague or suspiciously platitudinous.

In sum, this reviewer has doubts. But it is fair to say that the doubts are partly inspired by the value of other volumes in the same series under Dr Stark's general editorship. Inevitably one compares. And who would have Scheler when they can have Aristotle and Aquinas commentating or, for that matter, the Machiavelli of the *Discourses*?

J. WYNN REEVES

Migration and Mental Disease. By BENJAMIN MALZBERG and EVERETT S. LEE. New York: Social Science Research Council. 1956. Pp. x + 142. $1.50.

The authors of this book study the relationship between migration and mental disease by means of a statistical analysis of records of first admissions to hospitals for mental disease in New York State during the three years 1939–41. In all, 46,000 cases are included, classified into three main groups: dementia praecox, manic-depressive psychoses and all other psychoses. Migrants are variously defined as (*a*) foreign-born, (*b*) having been born in another state, (*c*) having migrated within the five-year period 1935–40. Rates of first admissions in the area of destination for migrants thus defined are compared with rates of first admissions for the 'indigenous' population of New York State. Age-standardized rates are used in comparisons. The authors are careful to point out many limitations in their data, and the danger of drawing unwarranted conclusions.

In general, migrants are found to show higher rates of first admission than non-migrants, whichever definition of migration is used. It is inferred that this differential holds for cerebral arteriosclerosis and senile psychoses as well as for the 'functional' psychoses. For foreign-born men the rate is relatively high for dementia praecox but not for other psychoses, while for foreign-born women the rates for all groups of psychoses are relatively high. For inter-state male migrants, the rate for dementia praecox is again high, while for inter-state female migrants the rate for manic-depressive psychoses is also high. Women show larger overall differentials than men. With the exception of dementia praecox, the rates for native (inter-state) migrants are higher than for foreign-born. Rates are much higher for recent arrivals.

Although the data do not enable one to decide between the two most plausible explanations of these differential rates—that they are due to initial selection, or to the problems of adjustment after migration—one finding appears to have some bearing on the hypothesis that schizophrenics tend to migrate more readily. The rate of admission for dementia praecox is low relative to rates for manic-depressive and other psychoses during the first year after migration.

The book has an introduction of some forty pages by Dorothy Thomas, which provides a compact and valuable account of previous work on the problem.

D. GRAHAM

Crime and Social Action. By GEORGE GODWIN. London: Watts and Co. 1956. Pp. vi + 277. 18*s*.

This book sets out to gather together various contributions to criminology made over two centuries from widely diverse fields.

Part I examines the nature of the criminal on the basis of a somewhat arbitrary selection of investigators ranging from Lombroso to Wertham as the representative of the modern psychiatric approach. This section ends with an emphasis on heredity as the paramount factor in producing the 'criminal diathesis', the criminal being merely the inferior man. Part II is an account of the development of penal methods from Beccaria, through such English reformers as Fry and Howard, to a somewhat unfair assessment of the work of East and Hubert as a contribution to the development of current methods of psychological treatment for offenders. Part III, on 'Murder and Madness', consists essentially of a discussion on how far the M'Naghten rules are out of date as a basis for dealing with problems of insanity and criminal responsibility. To this is appended a somewhat tendentious chapter on the death penalty with special reference to such topics as the Evans case of recent years. Part IV is a light-weight account of detection and the forensic 'sciences'.

It is too much to accept the claim that all this constitutes 'a very comprehensive and well-documented study of social attitudes to the criminal' with the author's method transforming 'what might otherwise have been a book for the specialist into one with a broad appeal to the general reader'. Mr Godwin's technique is a mixture of popular journalistic reporting, dramatic historical reconstruction, *Readers' Digest* condensation of information and ideas and quotations from original sources. The general presentation of all this material is muddled and poorly integrated, with a number of inconsistencies and loose ends which remain after his final humanitarian-scientific summing-up. A preference for writing in short, chopped-up paragraphs tends to add to the book's general lack of cohesiveness.

Though specialists will find most of the book flimsy, its general emotional impact on the non-technical general reader will probably be to leave him with a vaguely more 'enlightened' understanding of the problems of crime and criminals.

R. L. MORRISON

VOLUME 48 MAY 1957 PART 2

INDIVIDUAL DIFFERENCES IN REACTING TO DELAYED AUDITORY FEED-BACK

By J. T. BEAUMONT AND B. M. FOSS
Department of Psychology, Birkbeck College

The speech of normal subjects is impaired when the subject hears himself speak a fraction of a second after uttering, that is, when auditory feed-back is delayed. A study was made of individual differences in showing this impairment. It was found that those showing the impairment in marked fashion were likely to have high scores in tests of other kinds of 'perseveration'.

I. INTRODUCTION

It has been known for some time (Lee, 1951) that delayed auditory feed-back impairs the speech of normal subjects. There are four major effects: (1) a general slowing down of speech; (2) a prolonging of syllables—'illuminated' might become 'illooomminated'; (3) a repetition of consonants—'twittering' might become 'twittittering'; (4) occasional complete stops or blocking. The induced stutter shown in effect (3) generally occurs when the subject attempts to maintain a natural speed of talking. Effect (4) was described by one subject as 'allowing my ears to catch up with my voice'. There are two main ways of avoiding the stutter, namely, by speaking very slowly, and by speaking in short, rapid bursts. Both methods result in an overall slowing down of speech.

Individual differences in susceptibility to the artificial stutter are so marked and consistent that it seems reasonable to try to relate these individual differences to other characteristics of the subject. In choosing the tests used for determining these other characteristics, two 'hunches' were involved. The first was that individuals might differ in their ability to ignore feed-back, or in their ability to use it in a novel way; this might or might not be independent of the modality in which the feed-back occurred. The second was that the delayed feed-back interference is related to perseveration. It might be that susceptible individuals would also be susceptible to other kinds of perseveration.

II. EXPERIMENTAL

(i) *Apparatus and tests.* The delayed auditory feed-back was supplied by a modified tape recorder. The subject wore headphones and talked into a microphone. His voice was recorded on the tape and relayed back to him from a pick-up head which was a small distance further along the tape than the recording head. The delay could be adjusted by altering the speed of the tape. This was adjusted to give a delay of approximately one-fifth of a second. The other tests were as follows: (*a*) Star tracing. This was the familiar test in which the subject has to trace between the contours of a double contoured star while observing his movements in a mirror. (*b*) Mirror writing. The subject had to write letters and digits normally while observing his movements in a mirror so that the characters would be seen by the subject in mirror image form. (*c*) Einstellung test. This was devised by Luchins (1942). In it, a set was created to solve simple arithmetical problems in one, rather complex, way. Later, six critical problems were given in which this same method could be used for solution, or an alternative simpler method could be used. If subjects persisted in using the original method, they were said to show the Einstellung effect. (*d*) Imagery rigidity. This was a test invented by Gordon (1949). It consisted of a series of eleven questions involving the manipulation of the image of a car, e.g. 'what colour is it?', 'can you see it lying upside down?' Subjects with rigid imagery were said to find the manipulation of their images very difficult. That is, once having seen an image of a certain car in a certain situation, such subjects would be unable to see it any other way.

The star tracing and mirror drawing tests were included because they both involved the utilizing of visual feed-back in a novel way; also, they were tests of one kind of perseveration—habit perseveration. The Einstellung and imagery rigidity tests were also thought to test habit perseveration.

(ii) *Subjects and procedure.* The subjects were thirty non-psychologists with an age range of 18–30 years. The visual tests, followed by the Einstellung and rigidity tests, were carried out two weeks after the test with delayed auditory feed-back.

Delayed auditory feed-back: the subject was shown the tape-recorder and a brief account of its performance was given. He then put on the earphones and spoke a few words into the microphone. This familiarized him with the feel of the apparatus and also allowed the initial shock of the effect to pass off. After removing the earphones, he was given a page of printed matter of 400 words length and asked to read it aloud in his normal speaking voice. Shortly after he had finished, he was asked to recite aloud two short nursery rhymes twice each, the total length of utterance being ninety-six words. This performance was then repeated under experimental conditions, i.e. with the subject wearing earphones and hearing his voice $\frac{1}{5}$ sec. after he had spoken. During both control and experimental performances, the experimenter noted the time for each 50th word during the reading and the overall time only for the recitation. As a check of reliability, nine subjects were given a re-test on the reading, using a different passage, about two weeks after the first trial.

Manual-visual activity: both these tests were carried out at the same time some three weeks after the above test. Subjects were required to write letters in disconnected script and the digits 1–9 for periods of 20 sec. each alternately. After one period of each under normal conditions, a screen and mirror were placed so that the subject could observe his performance only in the mirror. Four alternations of letters and digits were completed under these conditions. The screen and mirror were then removed and a 5-point star placed before the subject. He was asked to trace its outlines as rapidly as possible without making any mistakes. This was then repeated with the screen and mirror in place: the usual procedure for mirror tracing. A single control and four experimental runs were timed, followed by four more alternations of letters and digits and finally, after removal of the screen and mirror, one more control period of each. The numbers of letters and digits written in each period and the times taken for star-tracing were recorded. The Luchins Einstellung test and the Gordon test were administered in the way described by their inventors.

III. RESULTS

Table 1 shows the mean scores for all except the Luchins and Gordon tests. The last column indicates the effect which the experimental conditions had on performance, either as percentage increase in time taken on a task, or as percentage decrease in amount of work done in a given time. No assumptions could be made about the distribution of scores, so subjects were placed in rank order for their performance on each test, and non-parametric techniques were used in assessing contingencies. Correlations were calculated with Kendall's tau.

(1) *Reliability*. A reliability test was made on the auditory feed-back experiment since nothing was known about consistency of performance on it. For the nine subjects used, the correlation was 0·83 with a standard error of 0·27, which was considered satisfactory.

Table 1. *Means and ranges of measures used on the visual and auditory tests. Thirty subjects*

	Control		Experimental		Effect	
	Mean	Range	Mean	Range	Mean	Range
		(in sec.)			(% increase)	
Reading						
First quarter	18·3	14–24	28·0	18–39	54·5	−4–155
Total task	155·5	130–198	221·7	155–368	47·1	−4–165
Star-tracing						
First star	15·1	9–26	207·5	42–639	1242·0	250–3790
Average for 4 stars	—	—	121·4	38–279	761·0	218–2150
Recitation	25·1	11–32	44·2	14–88	78·2	17–309
		(Amount written in given time)			(% decrease)	
Writing						
First trial	135·6	94–196	88·0	23–159	35·8	10–79
Average for all trials	—	—	192·5	67–349	29·4	4–70

(2) *Relationships between subtests*. There was a significant correlation of 0·33 between the overall effect on star-tracing and the overall effect on writing. In the auditory tests, there was a significant correlation of 0·56 between the overall effects on reading and reciting. The Gordon and Luchins tests did not correlate.

(3) *Adaptation*. In all the experiments in which feed-back was interfered with, there were significant correlations between the initial effect of the experimental condition and the speed of adapting to it. That is, the worse the impairment on the reading tests, the more rapid the adaptation. (The most likely explanation of this is that if subjects were little affected by the experimental conditions, they were working at something like their normal speed, so that little improvement could show; whereas if their performance was considerably affected, there would be greater opportunity for improvement.)

(4) *Relationships between the auditory and visual tests*. The overall effect on reading correlated 0·11 (not significant) with overall effect on writing, and 0·30 (significant) with overall effect on star-tracing. The overall effect on reciting was correlated positively, but not significantly, with the effects on the visual tests.

(5) *The Luchins Einstellung tests*. Table 2 shows the scores of subjects classified according to whether they showed complete Einstellung or not. The scores on other tests of these two groups were compared using the Mann-Whitney U test (one-tailed), the hypothesis being that those subjects showing complete Einstellung would be highly perseverant on the other tests. The probabilities, in the last column, show the significant contingencies that exist.

(6) *The Gordon imagery test*. This gave only one significant correlation, which remains unexplained. Subjects who said they were unable to visualize at all were quick reciters under normal conditions. Statistically this was a highly significant relationship.

Table 2

(Subjects are classed according to whether they showed complete Einstellung or complete lack of it. The 4 in-between subjects are omitted. The figures are in percentages, as in Table 1.)

	Einstellung ($N=11$)		Non-Einstellung ($N=15$)			
	Mean	Range	Mean	Range	U	p
Reading						
Effect on first quarter	69·6	14–152	44·0	−4–155	55	>0·05
Overall effect	62·0	24–165	25·2	−4–94	37	<0·01
Star-tracing						
Effect on first trial	1730	656–3790	876	250–1615	30	<0·01
Overall effect	967	305–2150	604	218–945	40	<0·05
Recitation	89·9	17–309	63·2	19–158	57	>0·05
Writing						
Effect on first trial	41·7	13–79	32·0	10–55	59	>0·05
Overall effect	33·7	5–70	26·0	4–50	63	>0·05

IV. DISCUSSION

The first 'hunch' tested was that subjects might differ in their ability to ignore feed-back, or to use it in a novel way. The low positive correlations between the auditory and visual effects might indicate that there were such consistent individual differences. However, the significant correlation between writing and star-tracing throws doubt on this, since, in writing, the mirror image has to be ignored, whereas in star-tracing it has to be used in a novel fashion. This suggests that the second 'hunch' is more plausible: that the tests are sampling individual differences in perseveration. The significant contingencies existing between results on the Einstellung test and the reading and writing tests add weight to this argument.

Eysenck (1947, pp. 154–6) has divided the various tests of perseveration into five classes, and suggests that the correlations between classes are low or negligible. The present study would suggest that two of these classes of perseveration are significantly correlated. They might be defined as follows (adapted from Lankes, 1914). First, a high perseverant is an individual whose performance on a task is more than normally impaired by the after effects of previous movements, or previous stimulation. (The stuttering machine—delayed auditory feed-back—assessed perseveration of this kind.) Secondly, a high perseverant is an individual whose performance on a task is more than normally impaired by previously learnt habits which conflict with the present task. (The Einstellung test, star-tracing and writing while looking in a mirror seem to be examples of this kind of perseveration. The writing test seems the least good measure, having smaller correlations, since the writing habit can be carried on successfully if the subject ignores what he sees in the mirror.)

It is interesting that Solomon (1952) discovered that natural stutterers showed more pronounced Einstellung than a normal group and that this agrees with common clinical experience that stutterers tend to have 'rigid' personalities. This suggests that people who are susceptible to delayed auditory feed-back may have something in common with natural stutterers. This is not so surprising in view of demonstrations by Cherry, Sayers & Marland (1955) that natural stutterers are able to talk at a more normal rate if their auditory feed-back is blocked by a loud masking noise.

The authors wish to express their great indebtedness to Dr D. M. Mackay of King's College, London, who lent them his apparatus for providing delayed auditory feed-back. Thanks are also due to the subjects who so kindly submitted to the indignities of the machine.

REFERENCES

CHERRY, E. C., SAYERS, B. McA. & MARLAND, P. M. (1955). Experiments on the complete suppression of stammering. *Nature, Lond.*, **176**, 874.

EYSENCK, H. J. (1947). *Dimensions of Personality*. London: Kegan Paul.

GORDON, R. (1949). An investigation into some of the factors that favour the formation of stereotyped images. *Brit. J. Psychol.* **39**, 156–67.

LANKES, W. (1914). Perseveration. *Brit. J. Psychol.* **7**, 387–419.

LEE, B. S. (1951). Artificial stutter. *J. Speech Dis.* **16**, 53–5.

LUCHINS, A. S. (1942). Mechanization in Problem Solving: the effect of Einstellung. *Psychol. Monogr.* **54**, no. 248.

SOLOMON, N. D. (1952). A comparison of rigidity of behaviour manifested by a group of stutterers compared with 'fluent' speakers in oral and other performances as measured by the Einstellung effect. *Speech. Monogr.* **19**, 198–9.

(*Manuscript received* 23 *September* 1955)

DISCRIMINATION AND OPERANT EXTINCTION

By H. M. B. HURWITZ and J. CUTTS
Birkbeck College, University of London

The experiment reported in this paper was performed to test the hypothesis that experimental extinction of a conditioned operant is facilitated when the stimulus difference between conditioning and extinction trials is increased. Previously this problem has been approached in two ways (Bersh, 1951; Dinsmoor, 1951, Notterman; 1951; Skinner, 1938; Bitterman, Feddersen & Tyler, 1953; Winnick, 1951).

First, extinction trials were differentiated from conditioning trials by omitting cues which had been an invariant feature of the reinforcement situation, e.g. the colour of the goal box in a maze-running experiment was changed. And, secondly, cues associated with the reinforced operant were omitted during extinction trials, e.g. the click of the feeding mechanism in the Skinner box was eliminated. Both experimental techniques are subject to the same limitation: eliminating or reducing cues previously associated with reward may induce a general decline in activity without necessarily affecting the rate of extinction. To overcome this objection we decided to investigate the part played by discrimination in conditioned operant extinction by adding additional cues to the familiar stimulus complex. Thus, whenever the conditioned operant occurred, it was accompanied by a novel stimulus. In all other respects the experimental situation was kept identical to the training trials. In the following experiment hungry rats were trained to press a bar in order to secure a small ration of food. Each bar response was accompanied by a flash of light. During extinction trials, the bar response was accompanied by light and a faint-sounding buzzer. (A pilot experiment, in which a group was trained with a buzzer sounding after each operant, and buzzer-light was presented during extinction trials, yielded roughly the same results as are reported here. We preferred using light as the training stimulus because it appeared to facilitate training. We have since found that light may be used as a reinforcer of operant responses in its own right. See Hurwitz (1956).)

In order to test the prediction that the additional stimulus facilitates operant extinction, it is necessary to rule out the alternative factors: general decline of activity and also specific decline of food-orientated behaviour. In a recently reported study of Melching (1954), statistically significant differences between experimental groups could not be accounted for satisfactorily, since the functional properties of the stimulus could not be identified. Melching's research design—which has many points of similarity with the present study—is outlined in Table 1.

Groups A and F may be regarded as control groups with respect to B and E. An analysis of the data showed that A *v.* B=1%, A *v.* E=1%, *A v. F*=5–10%, E *v.* F= 1%, *B v. E*=5%, *F v. B*=*not significant.*

Table 1

Groups	Frequency of buzzer in training (%)	Frequency of buzzer in extinction (%)	No. of extinction responses
A	100	100	755
B	100	0	574
E	0	100	402
F	0	0	637

The significance values between the various groups indicate that a simple stimulus generalization hypothesis will not account for all the results. Thus, the control groups, A and F, do not appear to be derived from a homogeneous population, and the experimental groups, B and E, reveal a difference amongst themselves depending on the kind of stimulus deviations from training conditions experienced.

In the present study, an attempt is made to obtain a measure of the strength of food-orientated behaviour, as distinct from critical operant responses. The food trough was fitted with a light plastic door. Opening the door closed a circuit and operated a counter. Any changes in food-oriented response vigour under either of the two experimental conditions could thus be detected. Our hypothesis (that the additional stimulus associated with the performance of the operant facilitates its extinction) would be supported if (*a*) the operant frequency under the two experimental conditions differed significantly, (*b*) the frequency of trough responses was constant for both situations.

Subjects. The subjects were twenty-two 3-month-old experimentally naïve male hooded rats drawn from the small inbred colony maintained by the Department of Psychology.

Apparatus. The apparatus consisted of a modified Skinner box, housed in a large, semi-sound-proof chamber. The subject could be observed through glass screens. As in a previously reported study (Hurwitz, 1955) the bar was placed 5 in. to the left of the food trough, thus forcing the subject to shuttle between bar and trough. The bar was mounted 3 in. above floor level. It measured ½ in. square and required a pressure of 2 g. to move it through ¼ in. The food trough was shielded by a light opaque door, which had to be nuzzled open before food could be retrieved. The electric circuits were so arranged as to deliver only one pellet of food at a time, i.e. it was impossible for the subject to accumulate a heap of food by repeated pressing on the bar. A small lamp-holder was mounted 7 in. above the bar. When the lamp was on, it gave 3·5 ft.-candles. The buzzer was placed 12 in. away from the box. It emitted a low buzz. On the wall, facing the bar and trough, we fitted a red indicator lamp which provided sufficient illumination for the subject to be observed through the glass screens surrounding the sound chamber. The food delivery machine used in this study has been described elsewhere (Hurwitz, 1955). Its principal merit is its almost noiseless operation.

Records. Records of the subjects' behaviour were obtained in two ways: (*a*) two magnetic counters recorded the number of trough and bar responses; (*b*) a three-channel pen recorder provided data on the serial ordering of responses and time intervals between responses.

Habituation and feeding schedule. Since the subjects were laboratory bred, they were accustomed to frequent handling and proved extremely tame. Approximately 2 weeks before the first training session, each subject was placed into a special feeding-fasting schedule apparatus which automatically made food available once daily for 2 hr. Two days before the first training trial, the subject was transferred from the feeding-fasting apparatus to a special cage, where it remained until the conclusion of the experiment. Each subject was fed once daily, at the time set aside for its experimental trials.

Training. Training extended over 3 days. On day 1, the subject was placed in the Skinner box and allowed to retrieve ten pellets from the food trough. Thereafter, any additional pellets supplied were accompanied by a brief (0·5 sec.) flash of light. This was repeated 20 times. The subject was then left to its own devices. However, whenever it moved towards the bar, E switched on the light. If the bar was actually touched a pellet of food was released into the trough. Ten pellets were delivered in this manner. Thereafter the subject was left in the box until it had secured ten further pellets.

On day 2, the subject was retained in the box until it had obtained thirty pellets by pressing the bar. On day 3, the subject was allowed to secure sixty pellets. If the subject took longer than 40 min. it was withdrawn from the experiment. Only one subject failed to satisfy this criterion of proficiency.

Experimental extinction. On the final day of the experiment each subject was allocated to one of two groups. For group I, the stimulus conditions associated with the critical operant were identical to the training phase: for group II, the stimulus conditions associated with the critical operant were different in that a new stimulus, the faint buzz, was added to the accustomed complex. Each subject was placed in the apparatus and allowed to obtain three pellets of food under the same conditions as marked training trials. Thereafter the experimental extinction session proper commenced. After 1 hr. the subject was removed from the apparatus and the experiment terminated.

Table 2. *Statistical comparison of difference among group I and group II with respect to bar and trough responses*

	Bar responses		Trough responses	
	Mean	S.D.	Mean	S.D.
Group I ($N=11$)	234·9	59·8	200·3	78·1
Group II ($N=11$)	188·5	37·8	229·0	59·0
	$t=2{\cdot}1$, $p=0{\cdot}05$		$t=0{\cdot}82$, p not significant	

Results. Tests of statistical significance were applied to the operant frequency scores and trough response scores for both groups. As $t=2{\cdot}1$ in the case of bar responses, and $p=0{\cdot}05$, the two groups may be said to differ significantly from each other. However, in the case of food-trough responses, $t=0{\cdot}82$, i.e. the differences were not significant.

These results accord with the requirements set out in the introduction for supporting the hypothesis that the additional stimulus associated with the operant during extinction trials facilitates extinction.

The project was financed by the Central Research Fund, University of London. The feeding-fasting schedule apparatus was constructed by Mr E. Wasservogel and Mr J. Brightman.

REFERENCES

BERSH, P. J. (1951). The influence of two variables upon the establishment of a secondary reinforcer for operant responses. *J. Exp. Psychol.* **41**, 62–73.

BITTERMAN, M. E., FEDDERSEN, W. E. & TYLER, D. W. (1953). Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.* **56**, 456–64.

DINSMOOR, J. A. (1951). The effect of periodic reinforcement of bar pressing in the presence of a discriminative stimulus. *J. Comp. Physiol. Psychol.* **44**, 354–61.

HURWITZ, H. M. B. (1955). Response elimination without performance. *Quart. J. Exp. Psychol.* **7**, 1–7.

HURWITZ, H. M. B. (1956). Conditioned responses in rats reinforced by light. *Brit. J. Anim. Behav.* **4**, 31–3.

MELCHING, W. H. (1954). The acquired reward value of an intermittently presented neutral stimulus. *J. Comp. Physiol. Psychol.* **47**, 370–74.

NOTTERMAN, J. M. (1951). A study of some relations among aperiodic reinforcement, discrimination training and secondary reinforcement. *J. Exp. Psychol.* **41**, 161–9.

SKINNER, B. F. (1938). *The Behavior of Organisms.* Appleton-Century-Crofts, New York.

WINNICK, W. A. (1951). Light discrimination based upon periodic reinforcement. *Amer. J. Psychol.* **54**, 573–9.

(*Manuscript received* 18 *December* 1955)

THE PSYCHOLOGICAL REFRACTORY PHASE*†

By THOMAS MARILL

Massachusetts Institute of Technology

A reaction-time experiment was performed in which *S* was presented with two lights and two hand-keys. A flash of the left light served as a stimulus to press the left key with the left hand; the right light to press the right key with the right hand. Trials were given in which both lights flashed; the interval between flashes being randomly selected from nine possible values ranging from 0 to 600 msec., the various intervals occurring with equal probability.

If s_1 and s_2 are the prior and subsequent of a pair of stimuli calling forth responses r_1 and r_2, respectively, it was found: (*a*) that presenting s_2 during the s_1–r_1 interval has no systematic effect on this interval; (*b*) that the s_2–r_2 interval varies markedly as a function of the s_1–s_2 interval, with s_2–r_2 longest for short s_1–s_2.

Starting in the 1880's research on the refractory phase has created a voluminous literature. In the early investigations 'refractory phase' was understood to apply to the difficulty in the rapid re-elicitation of not only nerve and muscle fibres but also of reflexes such as the eye-wink (Dodge, 1913) and the knee-jerk (Dodge, 1926), of voluntary responses (Telford, 1931), of associative (Thorndike, 1927) and of intellectual (Dodge, 1917) processes. In almost all of these cases the interest was in the *magnitude* of responses. More recent analyses of response mechanisms (Craik, 1947, 1948; Fitts, 1951; Hick, 1948; Vince, 1948; Welford, 1952) have made mention of a *psychological* refractory phase; the interest here has been primarily in the *latency* of responses upon rapid restimulation.

Very few experiments have dealt with this latter problem, the principal studies cited being those of Vince (1948) and Telford (1931). Vince's *S*'s were required to follow the motion of a small dot with the point of a pencil. The motion of the dot was discontinuous, consisting of steps to the right and left within a horizontal aperture. Occasionally two steps of opposite direction occurred in quick succession. Reaction-time (*RT*) in this situation was defined to be the interval between the step and the beginning of the appropriate tracking response. It was found that *RT* to the second of a pair of steps was markedly longer than normal if this second step came within 0·5 sec. of the first. (The following abbreviations will be used: s_1 and s_2 are the prior and subsequent of a pair of stimuli calling forth, respectively, responses r_1 and r_2. 'RT_1' and 'RT_2' are used interchangeably with, respectively, 's_1–r_1 interval' and 's_2–r_2 interval'. For the sake of brevity we also say, for example, 'the *RT to a stimulus*' to mean 'the interval between a stimulus and its appropriate response'.)

Telford investigated a situation in which *S*'s were instructed to press a telegraph key in response to an auditory stimulus. The stimuli were presented in a long sequence, the interval between any two being randomly chosen from a set of four possible intervals. It was found that the *RT* to a stimulus occurring after a 0·25 sec. interval was abnormally long.

---

* This research was supported in part by the United States Air Force under Contract No. AF 18 (600)–322, monitored by the Human Factors Operations Research Laboratories, Air Research and Development Command, Bolling Air Force Base, Washington 25, D.C. The writer is grateful to Dr J. C. R. Licklider, without whom this experiment could not have been performed.

† Since the present experiment was performed, and reported in abstract form (Marill, 1953), a similar experiment has been published by Elithorn & Lawrence (1955). The present data may be taken as supplementing those of Elithorn & Lawrence.

Discussions of the psychological refractory phase have assumed implicitly that the effect had its locus in the central nervous system. In the type of experiment discussed, however, the mechanics of limb movement obscure the facts, if any, of such a central refractoriness. From purely mechanical considerations we can expect it to take longer to initiate a hand-movement to the left when the hand is moving to the right than when the hand is stationary. This might account for part of the effect observed by Vince. Likewise in Telford's experiment we must consider the fact that at short stimulus-intervals the second key-tap of a pair was probably initiated at a time when the hand was just releasing the key after the first tap.

In view of these considerations it was felt to be desirable to perform an experiment on pairs of *RT*'s in which anatomically independent responses could be studied.

METHOD

Apparatus. (The apparatus described in this section made use of the following commercially available units: Stimulus lights: GE AR2 and GE NE36; oscillator: H-P Frequency Standard 100A; counters: Berkeley 410; pre-determined counter: two Grason-Stadler 461, in cascade.) A dual *RT* apparatus was employed. The *S* was presented with two lights and two hand-keys. In response to a flash of the left light (*L*-light) *S* pressed the left key with his left hand; in response to the *R*-light he pressed the right key with his right hand. Provision was made for measuring both *RT*'s independently and for flashing the two lights in quick succession in either order.

During the session *S* sat alone in a dimly lighted room. His forearms rested on a table in such a way that his hands could comfortably be kept on the keys. At the other end of the table, 5 ft. from his eyes and perpendicular to his line of sight, was mounted a black square, 1 ft. on a side, on which was affixed a small mark which *S* fixated during the trial. Located with centre $\frac{7}{8}$ in. on each side of the fixation mark was a circular opening $\frac{3}{4}$ in. in diameter, from which protruded the front of an electric bulb. The *L*-light was a split-disk-electrode argon bulb giving blue light, and the *R*-light was a similar neon bulb giving red light. Only half the electrode lighted during a flash, presenting *S* with a luminous semicircle. Flashes were produced by pulses approximately 10 msec. in duration and -40V. in amplitude, superimposed on a -40V. d.c. level.

During the session *S* wore earphones by means of which white noise could be presented. This served the dual purpose of signalling to *S* that a trial was in progress and of masking disturbing sounds from the laboratory.

The timers used to measure *RT* were constructed as follows: The output of a 10 kcyc. oscillator was fed through an on-off electronic gate into a high-speed counter. The gate opened as the stimulus flashed, letting the signal into the counter, and closed as *S* made his response, shutting off the input to the counter. The number of cycles of the 10 kcyc. signal occurring between the stimulus and the response was thereby displayed on the counter. Dividing this number by ten, we obtained the *RT* to the nearest millisecond.

The interval between flashes was generated by a predetermined counter connected through a gate to our oscillator. The gate opened simultaneously with the first stimulus.

After a predetermined number of cycles had passed into the predetermined counter, this unit generated an output pulse which flashed the other stimulus and started the second timer.

Routing switches were provided for flashing the two lights in either order and for flashing only one light. A separate circuit was also provided for indicating key-taps in the absence of any stimulus.

Subjects. Three male undergraduates at M.I.T., ignorant of the purpose of the experiment, were used as *S*'s.

Procedure. A 6 sec. period during which noise was presented to *S* defines the duration of a trial. During a session 108 trials were given at the rate of two each minute. These were broken up into three runs, between which *S* was allowed approximately 5 min. rest. All *S*'s took part in four training and nine experimental sessions. Each condition occurred the same number of times in all sessions, though in a different random order. The 108 trials of a session were broken down as follows:

(*a*) *Seventy-two binary trials.* In these trials both stimulus lights flashed. To make up such a trial a time interval between the two flashes was selected at random from nine possible intervals, ranging in 75 msec. steps from 0 to 600 msec. The interval between the onset of the noise and the first flash was selected at random from four possible intervals: 2·5, 3, 3·5, and 4 sec. One of the two possible orders of flashes was then also randomly selected.

(*b*) *Nine null trials.* It was desired to elicit as nearly as possible errorless performance from *S*. (An error would be said to occur if a key is tapped during a given trial without being preceded by a flash of its appropriate light or if a light flashes without being followed during the given trial by the appropriate key-tap.) In order to minimize the tendency to make errors, null trials were given. In these trials no light was flashed.

(*c*) *Twenty-seven simple disjunctive trials.* As a further check on errors and also to provide additional data, trials were given in which either the *R*- or *L*-light (but not both) flashed. In these trials the s_1–s_2 interval may be considered infinite, since s_2 never occurred.

In addition to the main experiment, as described above, a small subsidiary experiment was attempted, with the aim of finding the effect on *RT* of warning the subject in advance that a given one of the two lights would flash in the coming trial with a given probability. To make the two experiments comparable, eleven trials from this subsidiary experiment were randomly interspersed with trials of the main experiment during each session. The results from this subsidiary experiment proved inconclusive and will not be reported at this time. Since, in these subsidiary trials, *S* was always forewarned, and thus easily able to distinguish these trials from those of the main experiment, the probability that this addition affected the results of the main experiment is considered negligible.

RESULTS

The results of the experiment are displayed in Fig. 1. The 0-interval condition (two simultaneous stimuli) is plotted in both graphs, since the distinction between s_1 and s_2 cannot meaningfully be made in this situation. The single disjunctive trials are plotted in the left-hand graph over the designation '∞' (see (*c*) above). All points represent medians of samples of 72 scores with the following exceptions: the samples of 0-intervals contain 144 scores each, since we do not here distinguish between RT_1 and RT_2; the

Bureau Ednl. Psy. Research
DAVID HA INING COLLEGE
Dated
Accs. No

samples of the '∞' condition each contain 243 scores. The confidence-intervals were calculated according to the non-parametric method given by Mood (1950).

From the left-hand part of Fig. 1 we see that presenting s_2 during the s_1–r_1 interval has no appreciable systematic effect on this interval. Once initiated, RT's are not disturbed by the initiation of independent responses. If we consider RT_1 to be in some sense a 'normal' reaction time, then we see, from the right-hand part of Fig. 1, that a refractory phase is evidenced quite clearly only by subject 2. The comparison between RT_1 and RT_2 is not altogether satisfactory, however, for we must realize that the conditions under which RT_1 is made are not the same as those under which RT_2 is made. It is probably more meaningful to compare the RT_2 only among themselves. We see that there is a very significant change in RT_2 as a function of the stimulus interval. Our original conjecture, that such changes, as observed by Vince and Telford, were due to anatomical considerations, is thus seen to be false, since the responses in the present experiment were anatomically independent. It is quite clear that, in the present experiment, initiating a response during the early part of another reaction time leads to longer RT than initiating this response at a later time.

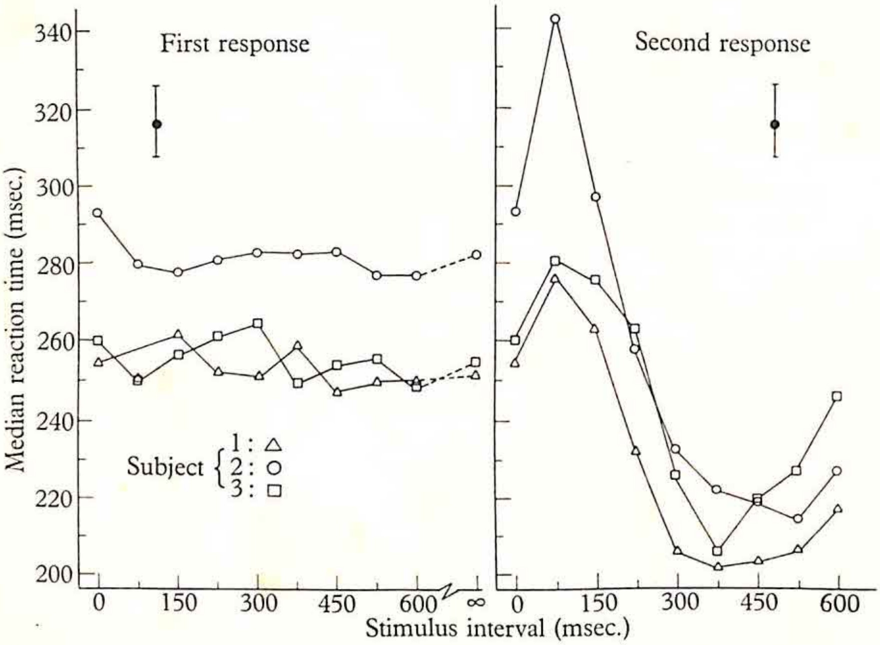


Fig. 1. Median reaction time for all subjects and conditions. The vertical bar in the upper corner of each graph indicates the size of the average 95% confidence interval for points in that graph.

REFERENCES

CRAIK, K. J. W. (1947). Theory of the human operator in control systems. I. The operator as an engineering system. *Brit. J. Psychol.* **38**, 56–61.

CRAIK, K. J. W. (1948). Theory of the human operator in control systems. II. Man as an element in control systems. *Brit. J. Psychol.* **38**, 142–8.

DODGE, R. (1913). The refractory phase of the protective-wink reflex. *Amer. J. Psychol.* **24**, 1–7.

DODGE, R. (1917). The laws of relative fatigue. *Psychol. Rev.* **24**, 89–113.

DODGE, R. (1926). Theories of inhibition. II. The refractory phase hypothesis of inhibition. *Psychol. Rev.* **33**, 167–87.

ELITHORN, A. & LAWRENCE, C. (1955). Central inhibition—some refractory observations. *Quart. J. Exp. Psychol.* **7**, 116–27.

FITTS, P. M. (1951). Engineering psychology and equipment design. In Stevens, S. S. (ed.): *Handbook of Experimental Psychology*, pp. 1287–1340. New York: Wiley.

HICK, W. E. (1948). The discontinuous functioning of the human operator in pursuit tasks. *Quart. J. Exp. Psychol.* **1**, 36–51.

MARILL, T. (1953). The psychological refractory phase (Abstr.). In *Final Report of Psychological Research Program for the Human Factors Operations Research Laboratories*. Research Laboratory of Electronics, Massachusetts Institute of Technology.

MOOD, A. M. (1950). *Introduction to the Theory of Statistics*. New York: McGraw-Hill.

TELFORD, C. W. (1931). The refractory phase of voluntary and associative responses. *J. Exp. Psychol.* **14**, 1–35.

THORNDIKE, E. L. (1927). The refractory phase in associative processes. *Psychol. Rev.* **34**, 234–6.

VINCE, M. A. (1948). The intermittency of control movements and the psychological refractory period. *Brit. J. Psychol.* **38**, 149–57.

WELFORD, A. T. (1952). The 'Psychological refractory period' and the timing of high-speed performance—a review and a theory. *Brit. J. Psychol.* **43**, 2–19.

(*Manuscript received* 26 *December* 1955)

THE NORMAL DISTRIBUTION OF INTELLIGENCE: A CRITIQUE

By D. G. LEWIS

Collegiate Faculty of Education, Aberystwyth

The dependence of the distribution of test scores on the facility and discriminatory values of the component items is discussed with special reference to the (approximately) normal distribution of quotients from the Binet Scale. A discussion follows as to how this should affect our interpretation of the 'normal distribution of intelligence'.

I. Introduction

The normal distribution occupies a place of central importance in psychological measurement. It is widely accepted that intelligence, along with other cognitive traits, is distributed according to the normal curve. There are many reasons for this. There is, for instance, the layman's misinterpretation of 'normal'. The term is associated with one of its senses in popular usage, and this in turn tends to conceal the supposition that any other distribution of intelligence would be abnormal in the sense of unnatural, a phenomenon to be explained. From this point of view it is unfortunate that 'normal' is generally preferred, as a label, to 'Gaussian', 'Laplacian' or 'probability integral'. To insist on the description of the normal distribution as one in which the frequency of any deviation from the mean is inversely proportional to the antilogarithm of half the square of that deviation (measured in units of standard deviation) might be somewhat repellent to the mathematically unsophisticated. It would serve as a salutary corrective, however, to the naïveté of popular interpretation. In particular, it would permit of no escape from recognizing that the 'normal' of normal distribution is a technical term (and in no way corresponds to the sense of the word in such statements as 'He does not enjoy normal health' or even 'He is of normal height'). It is by no means self-evident that a law of 'normality' underlies the distribution of human traits. Rather it is the existence of such a distribution, not departures from it, which requires explanation in the first place.

There is, also, the argument from analogy with other inherited characteristics. While 'intelligence', as a term in general usage, might not always be identified with innate ability, it is at least regarded as an attribute which is primarily innate. (The general usage would seem to correspond to the geneticist's 'phenotype' if not to his 'genotype'.) Many such attributes, of which height is perhaps the most frequently quoted example, are found to be distributed in a manner which adheres very closely to the normal distribution. 'Intelligence' then—except insofar as its distribution may be affected by environmental factors—can also, it is thought, be expected to exhibit this distribution. While this, too, is an error of laymen rather than psychologists there is a related reason which is more worthy of consideration. This is that the normal distribution of innate ability can be deduced, on certain assumptions, from genetic grounds.

Taking the genetic basis of intelligence as multi-factorial, each factor may be regarded as one of a pair of alternative or allelic genes. Thus in one such pair, Aa, only the allele A would contribute to intelligence, a being neutral. If the total number of pairs is large and each allele has always an equal chance of being present in the zygote then the distribution of the contributing factors, i.e. the *As*, would be normal, or approximately so. (This follows from the laws of chance in the same way that the successive tossing of a large number of coins would result in a distribution of 'heads' that is approximately normal.) Although this is undoubtedly 'an over-simplified genetic mechanism—no account, for instance, is taken of dominance, assortative mating and linkage—Fisher (1918) has shown that a much more general mechanism may be postulated and a normal, or near normal, distribution of the contributing factors still obtained' (personal communication from Sir Cyril Burt). Fisher chose stature as his illustration of an inherited trait, and Burt (1955) has suggested that the same reasoning would apply to intelligence defined as an innate ability. It is here that the distinction between intelligence so defined and intelligence as the level of cognitive efficiency actually attained becomes important (e.g. Hebb, 1949, ch. 11) and conclusions which seem justified with respect to the former concept may not be applicable to the latter.* Thus Maxwell (1955), in a discussion of the relative potency of heredity and environment on intelligence, contents himself with the modest claim that these genetical considerations imply 'some justification for the normal distribution of intelligence'.

Most important of all, however, is the claim that the normal distribution of intelligence has been established from the results of intelligence tests. In particular, this claim is advanced with respect to the individual tests of the Binet Scale (e.g. Burt, 1953). It is recognized that the normal distribution of quotients derived from Moray House tests is a consequence of the process of standardization (Thomson, 1932; Lawley, 1950). (The distribution of raw scores on these tests is usually rectangular rather than bell-shaped, and this is afterwards converted to 'normality'.) The distribution of quotients derived from the Binet Scale, on the other hand, is not a statistical artefact, and the empirical fact that these distributions have been found to agree fairly closely with the normal curve is often upheld as a convincing vindication of the normal distribution of intelligence. It will be shown, however, that distributions from both types of tests depend upon the principles governing the selection of test items. This must inevitably affect our interpretation of any resulting distribution of quotients.

II. Dependence on item selection

Two statistical criteria determine the selection of test items for Moray House tests (see Wilson, 1954). One is an index of item difficulty and the other an index of item discrimination. The former, nowadays termed the *facility value*, is merely the proportion of testees who pass the item. The latter indicates to what extent those who pass are the successful testees of the test as a whole. Many indices of item discrimination are possible. One is the E_{13} value and is determined by the numbers of those passing the item who are in the top and bottom thirds of the overall order of merit. It is

$$E_{13}=\frac{S_1 \overset{\frown}{\times} S_3}{\frac{1}{3}n},$$

where S_1 is the number of top-third successes, where S_3 is the number of bottom-third successes and n is the number of testees. Although theoretically E_{13} could vary between

---

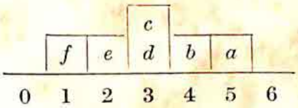
* Burt (1955) however estimates that less than 25 per cent of the variance of intelligence 'as an observable and empirically measurable trait' is due to environmental conditions.

+1 and −1, in practice its value is usually positive. Items whose E_{13} values are less than 0·5 are usually rejected as unsuitable for selection for a Moray House test (*op. cit.* p. iv). If a choice of suitable items is available values near the maximum +1 are favoured. The facility value of all accepted items must be near 0·5.

Only one statistical criterion determined the selection of test items for the 1908 Binet-Simon Scale, i.e. their facility values with respect to the particular age groups. Binet considered an item suitable for allocation to a certain age level if it could be passed by about two-thirds of the testees of that age. Burt (1921) in his adaptation of the scale, adopted instead the ratio one-half. This ratio is also used in the Stanford and Terman-Merrill Revisions (Terman, 1916, p. 53; Terman & Merrill, 1937, p. 9) and, in fact, still constitutes one of the main criteria for item selection. Another criterion is that the test items should differentiate between the various age-levels, i.e., the percentage of successful testees should rise steeply with age (Terman & Merrill, 1937). Thus as a child is taken through tests from the various age levels, from the basal age to the level at which all tests are failed, tests with a wide range of facility values are encountered. Item discrimination was not considered at all in the original construction of the scale. In the Stanford Revision, however, test items were examined from this point of view (Terman, 1916, p. 55) and from data subsequently supplied (Terman *et al.* 1917, p. 133) it would seem that the E_{13} values of most items are about 0·5, or a little less. Item discrimination is thus lower than is the case with group tests of the Moray House type. It may be shown that these differences in item selection can account for the contrasting types of frequency distributions obtained.

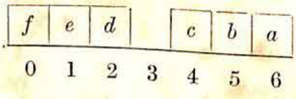
Consider a 6-item test undertaken by 6 (equal-sized) groups of testees—*a*, *b*, *c*, *d*, *e* and *f*. Suppose that in each of the following illustrations every member of any one group records the same response to each item. Suppose also that the order of merit of the groups, as given by the complete test, is alphabetical (except where two or more groups are equally placed). The E_{13} value of any item is then determined by the presence, in the item successes, of any of the four groups *a* and *b* (for S_1) and *e* and *f* (for S_3). This, together with the facility value (*F.v.*), is shown underneath each item. The resulting frequency distribution is shown alongside.

(i) Item no.	1	2	3	4	5	6
Successful groups	*a*	*a*	*a*	*a*	*a*	*b*
	b	*b*	*b*	*c*	*c*	*d*
	c	*d*	*e*	*d*	*e*	*f*
F.v.	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
E_{13}	1	1	$\frac{1}{2}$	$\frac{1}{2}$	0	0



Here the facility value of each item is $\frac{1}{2}$, the ideal for allocation to a Moray House test. The E_{13} values, however, vary more than would be the case for actually selected items.

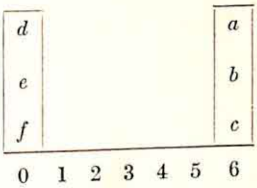
(ii) Item no.	1	2	3	4	5	6
Successful groups	*a*	*a*	*a*	*a*	*a*	*a*
	b	*b*	*b*	*b*	*b*	*d*
	c	*c*	*c*	*c*	*d*	*e*
F.v.	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
E_{13}	1	1	1	1	1	0



With a more strict selection with respect to the E_{13} values, the central concentration of (i) disappears and the tendency towards rectangularity, rather than normality, is evident.

(iii)

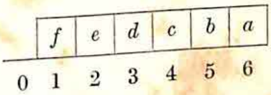
Item no.	1	2	3	4	5	6
Successful groups	*a* *b* *c*	*a* *b* *c*	*a* *b* *c*	*a* *b* *c*	*a* *b* *c*	*a* *b* *c*
F.v.	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
E_{13}	1	1	1	1	1	1



Taking an extreme case with each item having an E_{13} value of 1 the departure from normality is even more pronounced. This is the theoretical limit where all items are perfectly discriminating, each reproducing the result of the entire test. In practice its attainment would be impossible (and indeed undesirable!). The tendency for the distribution to be rectangular or U-shaped rather than bell-shaped is clearly illustrated, however, as items of low E_{13} values are rejected.

(iv)

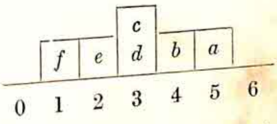
Item no.	1	2	3	4	5	6
Successful groups	*a*	*a* *b*	*a* *b* *c*	*a* *b* *c* *d*	*a* *b* *c* *d* *e*	*a* *b* *c* *d* *e* *f*
F.v.	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{5}{6}$	1
E_{13}	$\frac{1}{2}$	1	1	1	$\frac{1}{2}$	0



The facility values of the items now vary widely as is the case with tests from the Binet Scale. The extremity concentrations of (iii) disappear to yield a rectangular distribution.

(v)

Item no.	1	2	3	4	5	6
Successful groups	*b*	*a* *c*	*a* *b* *e*	*a* *c* *d*	*a* *b* *d* *f*	*a* *b* *c* *d* *e*
F.v.						
E_{13}	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$



Finally in an illustration in which an E_{13} value of $\frac{1}{2}$ is taken throughout and in which the facility value varies between $\frac{1}{6}$ and $\frac{5}{6}$ the central concentration of (i) returns.

It has thus been illustrated that the frequency distribution of tests is dependent on both the difficulty and discrimination of the component items. In particular: (1) when the facility value of all items is the same, or nearly so, the distribution tends to be rectangular rather than bell-shaped, especially if the items are highly discriminating (if all the items are highly discriminating the distribution might even be U-shaped); (2) when the facility value varies considerably and the items are moderately discriminating a distribution of central concentration is to be expected.

III. Discussion and conclusions

Many papers have appeared subjecting the suppositions of mental measurement to a critical examination. It is as well that this is so. The attempt to extend the precision of the so-called exact sciences to psychology must be scrutinized with care. Indeed some would maintain that the attempt is futile and that Whitehead's dictum 'to talk sense is to talk in terms of quantities' is inapplicable in the psychological field. Thus Chambers (1943) doubts whether the concept of measurement is applicable at all to psychological qualities and, more recently, Zangwill (1950, chap. 6) has echoed these views suggesting that nothing psychological is truly measurable. Similarly, Richmond (1953) emphasizes the distinction between assessment and appraisal, with the applicability of psychometric techniques confined to the former. However, measurement in psychology need not be identified with measurement in physics* (see Ehrenberg, 1955). Also it need not presuppose a *ratio* scale, where the existence of both a zero point and units of constant size is established. Measurement is also possible with an *interval* scale, e.g. the Fahrenheit scale of temperature, which has no true zero and thus does not permit the comparison of ratios (so that, for example, a temperature of 100° F. does not mean twice as hot as 50° F.). Of course the comparison of differences between successive points on the scale would still be permissible. If, however, the units were not constant in size even this comparison would be vitiated. Such a scale would then be an *ordinal* scale, and would provide a ranking of qualities with no available information concerning possible differences between the ranks.

If measurement in psychology is regarded as essentially measurement on an ordinal scale much of the criticism would be assuaged. Thus while, on such a view, an I.Q. of 120 would denote greater intelligence than an I.Q. of 110, and an I.Q. of 110 greater intelligence than one of 100, it would be illegitimate to equate the differences between the two pairs of quotients. Intelligence would thus be measurable in the same way as, say, the hardness of minerals is measurable, and no attempt could be made to compare differences along the scale. Indeed, as Warburton (1955) points out, the applied psychologist would be content with no greater claim. He is interested primarily in classifying people and is concerned with rank order rather than absolutes. Not all psychologists, however, are content with this and it is thus claimed that intelligence quotients should be interpreted as measurements on an interval scale (e.g. Banks & Burt, 1953, p. 161), so that differences between quotients may be legitimately compared. A more ambitious claim still is that with a scaling technique such as Thurstone's (1925), intelligence measurement may be interpreted as measurement on a ratio scale. The basis of these interpretations is the normal distribution of intelligence. Quotients are interpreted as measurements on an interval scale only when their distribution for a complete year group is normal. Thurstone's method of formulating a ratio scale is based on the assumption that this is so for successive year groups. A more recent attempt by Peel (1949) to construct a scale with constant and additive units is also dependent on this assumption.

The approximately normal distribution of Binet quotients is, as has been stated, an empirical fact. The normal distribution of quotients from Moray House tests, on the other hand, is imposed on the obtained scores during the process of standardization. It could be argued that this imposition is more than a statistical artefact in that it enables measurements by these tests, which have been constructed via the statistical criteria of item selection described, to conform with those of the Binet tests. However, results from the Binet Scale are also dependent on item selection. Its bell-shaped distribution of quotients, no less than the roughly rectangular distribution of Moray House tests, may be regarded as a consequence of the facility and E_{13} values of the component items. It would therefore be imprudent to regard the near normal distribution of Binet quotients as verification of a brute fact concerning the distribution of intelligence. In much the same way as the existence of an absolute zero of temperature is not so much an experimental fact about thermal phenomena but, rather, a conceptual matter—a consequence of the way in which degrees of warmth are put into relation with a number series (see Toulmin, 1953, pp. 129–133)—so, too, the distribution of test quotients (and in particular of Binet quotients) is a consequence of the way in which the test has been constructed. This need not, of course, vitiate the standardization of test quotients to a normal distribution, nor even their subsequent interpretation as measurements on an interval scale—provided it is recognized that such measurements are dependent upon the kind of measuring instrument employed.

---

* Thus the physicist Campbell has maintained that the *operation* of measurement is essentially the physical act of placing end to end (Campbell, N. R., *Physics: The Elements* (1920), chap. 10). This excludes a comparable operation in psychology. The *concept* of measurement, on the other hand, may be considered that of comparing two or more classes (of objects, individuals, etc.) with respect to a common abstracted quality—and this would be applicable in psychology no less than in physics.

The paper commenced with a critical appraisal of some of the reasons for accepting the normal distribution of intelligence as a fundamental fact of psychology. One further reason will be adduced in conclusion. The distribution agrees with common-sense notions as to how native ability is distributed. We all recognize that the number of geniuses is relatively small. So, too, is the number of mental defectives. Most persons are average, or about average, in ability. The normal distribution of intelligence fits in with this. Of course what is really necessitated from common sense is not the normal distribution as such, but a humped or bell-shaped distribution of which the normal distribution is but an instance. Why then should it be the normal distribution which is taken as the best representation of fact? The answer, perhaps, is that owing to the extreme usefulness of the distribution in the statistical theory of sampling so many of its properties have been investigated. As I was once told by a colleague, a specialist in test construction, it is convenient to have intelligence follow the normal curve; we know so much about it.

REFERENCES

Banks, C. & Burt, C. (1953). Statistical analysis in educational psychology. Chap. 13 of *Current Trends in British Psychology*. (Edited by Mace, C. A. & Vernon, P. E.) London: Methuen.

Burt, C. (1921). *Mental and Scholastic Tests*. London: Staples.

Burt, C. (1953). *Times Educational Supplement*, No. 2009 (30 October), Correspondence Column, p. 925.

Burt, C. (1955). The evidence for the concept of intelligence. *Brit. J. Educ. Psychol.* **25**, 158–77.

Chambers, E. G. (1943). Statistics in psychology and the limitations of the test method. *Brit. J. Psychol.* **33**, 189–99.

Ehrenberg, A. S. C. (1955). Measurement and mathematics in psychology. *Brit. J. Psychol.* **46**, 20–9.

Fisher, R. A. (1918). The correlation between relations on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinb.* **52**, 399–434.

Hebb, D. O. (1949). *The Organization of Behaviour*. London: Chapman and Hall.

Lawley, D. N. (1950). A method of standardizing group-tests. *Brit. J. Psychol. Statist. Sect.* **3**, 86–9.

Maxwell, J. (1955). Heredity, environment and intelligence. *Times Science Rev.* no. 15, 8–9.

Peel, E. A. (1949). Item difficulty as the measuring device in objective mental tests. *Brit. J. Statist. Sect.* **2**, 69–75.

Richmond, W. K. (1953). Educational measurement: its scope and limitations: a critique. *Brit. J. Psychol.* **44**, 221–31.

Terman, L. M. (1916). *The Measurement of Intelligence*. London: Harrap.

Terman, L. M. & Merrill, M. A. (1937). *Measuring Intelligence*. London: Harrap.

TERMAN, L. M. *et al.* (1917). *The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence.* Educ. Psychol. Monographs, no. 18.

THOMSON, G. H. (1932). The standardisation of group tests and the scatter of intelligence quotients. *Brit. J. Educ. Psychol.* **2**, 92–112, 125–37.

THURSTONE, L. L. (1925). A method of scaling psychological and educational tests. *J. Educ. Psychol.* **16**, 433–51.

TOULMIN, S. (1953). *The Philosophy of Science.* London: Hutchinson's University Library.

WARBURTON, F. W. (1955). The scientific status of mental measurement. *Brit. J. Psychol.* **46**, 122–9.

WILSON, E. B. (1954). The Construction of a 13+ Group Intelligence Test. Unpublished B.Ed. Thesis, University of Edinburgh.

ZANGWILL, O. L. (1950). *An Introduction to Modern Psychology.* London: Methuen.

(*Manuscript received* 18 *January* 1956)

A ONE-WAY WINDOW

By C. B. HINDLEY and FRANK FALKNER
The Child Study Centre, University of London Institutes of Education and Child Health

The wire mesh type of screen described by Gesell suffers from the disadvantage that absolute silence is called for on the part of any observers. Furthermore, it requires a big difference in illumination on the two sides of the screen, as does dark blue glass.

Glass, mirrored in strips, is expensive, and again requires a big difference in illumination.

Pinneau (1951) has described a way of making a partial mirror, which differs from an ordinary mirror in having a thinner layer of silver, and therefore permits the penetration of sufficient light for one-way viewing. However, this is troublesome to make, and expensive to obtain commercially.

The screen described is very easy to make, and has proved to be eminently satisfactory. It consists of a sheet of fine-gauge copper-wire mesh (30 to the inch), sandwiched between ordinary plate glass on the front surface (i.e. the testing room side), and dark blue glass, of the type used in ambulances, behind (i.e. the observation room side). The layers of the 'sandwich' are fixed together at the edges by transparent cellophane, which is entirely hidden within the groove of the window support.

Panels of this type can be inserted in the wall between an observation room and a testing room. From the testing room the appearance is of copper coloured panels, which children take to be part of the wall decoration.

As with the Gesell screen, the shiny surface of the wire mesh reflects a good deal of light and produces a diffuse dazzle on the front surface. The amount of light coming from the observation room is diminished by the blue glass. From the observation room side there is clear vision of the testing room, presumably because the blue glass, in absorbing much of the light, prevents dazzle from this surface of the wire mesh.

The window has proved to be satisfactory, even when the level of illumination is the same in the two rooms, though, of course, clarity is increased by heightened illumination in the testing room. Little sound penetrates the screen.

References

Gesell, A. (undated). *The First Five Years of Life*. London: Methuen.
Pinneau, S. R. (1951). A technique for making one-way mirrors. *Child Develpm.* **22**, 235–41.

CONFLICT AND CHOICE TIME

By D. E. BERLYNE*

Psychology Department, University of Aberdeen

Four variables have been suggested as determinants of degree of conflict: (1) the number, (2) the nearness to equality in strength, (3) the absolute strength, and (4) the degree of incompatibility, of competing response tendencies. The first two, but not the others, have figured in investigations relating reaction time to information theory. Three experiments were carried out to test the effects of all these variables on choice time. Two types of choice RT were recorded: first, the forced-choice RT (i.e. the traditional choice RT), for which one stimulus was presented at a time, each stimulus having its appropriate response; and second, the free-choice RT, for which two or more stimuli were presented together, S having to respond to any one of them.

The principal findings were as follows. (1) Both types of choice RT were longer when the number of alternative responses was increased. (2) Free-choice RT's were invariably longer than forced-choice RT's, confirming the influence of equality of strength. (3) Forced-choice RT's were inversely related to stimulus intensity, which was regarded as a factor influencing the absolute strength of response tendencies. Free-choice RT's were shortest at the intermediate of the three intensity levels studied and significantly longer at the highest and lowest levels. The proportion by which free-choice RT's exceeded forced-choice RT's increased with stimulus intensity. (4) Degree of physiological incompatibility did not affect free-choice RT's, although physiologically incompatible responses produced longer forced-choice RT's. (5) There were no significant differences between the RT's of S's with high and low neuroticism scores.

The expediency of continuing to include the degree of physiological incompatibility among the determinants of degree of conflict is called in question by the results. Otherwise, choice time, apart from the time taken up by eye movements, etc., which may be shorter with more intense stimuli, appears to increase with degree of conflict. The results are discussed in the light of various theoretical models, derived from Hull's principle of behavioural oscillation, Miller's theory of conflict, information theory, and Hebb's theory of cortical disruption.

I. INTRODUCTION

In the early stages of psychology, a great deal of time has necessarily had to be spent on artificially simple situations, in which an experimenter's attention can be focused on one response at a time. But more intricate situations, such as face both human beings and animals in their natural environments, raise a mass of new problems relating to interaction between simultaneously aroused response tendencies. One form this interaction can take is *conflict* (or *competition*), which has already proved to be important in such varied areas as abnormal psychology (Freud, 1929; Pavlov, 1927), perception (Bruner & Postman, 1949), emotion (Hebb, 1949), remembering (Gibson, 1940), and curiosity and thinking (Berlyne, 1954*a*, *b*). The possibility of introducing a variable representing *degree* or *severity of conflict* would therefore seem to be worth exploring. It has, in fact, been suggested (Brown & Farber, 1951; Berlyne, 1954*a*) that such a variable might profitably be regarded as an increasing function of four antecedent variables, namely, (1) the number of competing response tendencies, (2) their absolute strengths, (3) the degree to which their strengths approach equality, and (4) their degree of incompatibility or antagonism.

---

* Now temporarily at the Centre for Advanced Study in the Behavioural Sciences, Stanford, California, U.S.A.

One effect that has often been attributed to conflict is a *lengthening of reaction time or decision time.* It has been reported to occur when children have to choose between unpleasant rather than pleasant drinks (Barker, 1942) and when there is a conflict between what is expected and what is perceived (Desai, 1939; Wilcocks, 1928). A long association time is widely recognized as an indication that a word arouses emotional conflict (Jung, 1918). Choice reaction times are known to increase with similarity between the alternative stimuli, whether in some physical characteristic (Woodworth, 1938, pp. 333–4) or in affective value (Shipley, Coffin & Hadsell, 1945), and Dember (1954) has proposed a method of inferring 'stimulus distance' from decision time. Discrimination of similar stimuli may be taken to involve conflict, on the grounds that each stimulus tends, by generalization, to evoke the responses appropriate to the other stimuli as well as that appropriate to itself.

The theory of conflict links up in two ways with psychological applications of information theory. First, the information theorist's expression for 'uncertainty' or 'entropy', $-\sum_i p_i \log_2 p_i$, resembles degree of conflict in two respects, if we bear in mind the common acceptance of probability as a measure of response strength: it increases with the number of alternatives, and it reaches a maximum when their probabilities are equal. Secondly, choice reaction time has been shown, in certain conditions, to be a linear function of 'uncertainty' and, in particular, to increase as the alternatives increase in number and as they approach equiprobability (Crossman, 1953; Hick, 1952; Hyman, 1953; Klemmer & Muller, 1953). If we reinterpret 'number of alternative stimuli' to mean 'number of competing response tendencies, including those due to generalization', and 'equiprobability of alternative stimuli' to imply 'equal strength of competing response tendencies', we can see that reaction time has been shown to depend on two of the variables held to underlie degree of conflict. The experiments to be reported in this article were intended to investigate the relations between reaction time and all four suggested determinants of degree of conflict, including the two which have so far not been handled by information theorists. It was hoped that such an investigation might help both to integrate applications of information theory with the main body of psychology and to explore the usefulness of reaction time as a convenient measure of conflict, e.g. for the study of thought processes.

The experiments enabled the traditional choice reaction time (which we may call '*forced-choice RT*') to be compared with what we may call '*free-choice RT*'. In the forced-choice situation, only one stimulus is presented at any one time, and only the response associated with that particular stimulus is permitted. Response tendencies associated with the other stimuli may also occur by generalization, so that conflict between them and the correct response may explain why choice RT's are longer than simple RT's. But the correct response tendency will be considerably stronger than the others, so that a relatively slight degree of conflict will be entailed. In the free-choice situation, two or more stimuli are presented together. Each has its corresponding response, and S is left free to respond to any one of them. Two or more response tendencies of about equal strength will thus arise at once, and, since no more than one can be overtly expressed, a greater degree of conflict should result than with forced-choice reactions. We can predict, therefore, that free-choice RT's will exceed forced-choice RT's. It may be worth noting that Baker & Elliott (1948) have found this to be so in the field of verbal behaviour, since their experiment showed free-association latencies to be greater than controlled-association latencies. The free-choice situation can be seen to typify a large class of everyday dilemmas, in which a decision has to be made between alternative courses of action or as we say, several aspects of the environment 'compete for attention' (Berlyne, 1950, 1951*a*, *b*, 1957).

Our forced-choice situation, like most of the situations handled by psychological information theory or by psychophysics, demands what Platt (1956) calls a *polar decision*. Each of a set of alternative environmental events has its corresponding 'correct' response, so that the organism, in order to select its response, has merely to absorb enough information to identify the event. Platt points out that higher animals are often faced with the rather different problem of *equivocal decisions*, imposed on them when one response is not clearly favoured over others, because information, e.g. about remote consequences, is insufficient. Equivocal decisions are generally more time-consuming than polar decisions, and the effects of indecision may be biologically harmful both by wasting time and by generating 'anxiety'. The indecision may be resolved by converting an equivocal decision into a polar one, e.g. by some superstitious procedure, or by resorting to techniques which provide new information (Berlyne, 1954*a*). Our free choices represent yet a third sort of decision, not uncommon in everyday life, which we may call the *arbitrary decision*. The threat of indecision is here greater still, as one response is quite as good as another and receipt of further information cannot help.

Of the two variables whose effects on both kinds of choice RT were to be studied, the *absolute strength* (reaction potential) of response tendencies could be varied by varying the intensity of the stimuli evoking them (Hull, 1949; Logan, 1954). As for the *degree of incompatibility* between responses, a pair of movements that could be performed together —forward movements with both hands—was compared with a pair of movements that could not—forward and backward movements with the right hand. Both conflict theory and information theory have so far confined themselves to cases of complete incompatibility between responses. There is, however, evidence (Konorski, 1948, pp. 120 ff.) that certain responses, if aroused simultaneously, will weaken each other without complete mutual exclusion. It should be pointed out, of course, that any two actions which are *physiologically* capable of simultaneous performance acquire a *learned* incompatibility as a result of instructions not to perform both. Our experiments were therefore comparing a pair of alternative responses having learned incompatibility with another pair having both learned and innate incompatibility.

II. EXPERIMENT I

Apparatus comprised three units. First, a *stimulus unit*, comprising a screen with four $1\frac{1}{2}$ in. square apertures in a horizontal line at about eye-level. The apertures were covered with white paper, and small circular spots of light could be made to appear in the middle of any of them. A Palmer Timing Unit caused these visual stimuli to appear for 5 sec. with intervals of 5 sec. between them. E, sitting behind the screen, used push-buttons during the intervals to determine the colours and positions of the stimuli that were to appear next. S sat about 6 ft. from the board, and the room was dimly lit through chinks between black-out curtains and the edges of the windows.

Secondly, *two response units*. One, the C (or 'physiologically compatible') unit, consisted of two switches, each of which could be pushed forward only, mounted on a board. The other, the I (or 'physiologically incompatible') unit, consisted of one switch which could be pushed backward or forward, mounted on a different board.

Thirdly, a *Chronotron Millisecond Meter*, set to record RT's in units of 10 msec.

Subjects were twenty male junior-secondary-school pupils, all aged 12 years.

Design. Two sequences of 102 stimulus patterns were drawn up with the help of a random-number table. Each comprised forty-one *red* and forty-one *green* stimuli (for forced-choice RT's) and, mingled with them, twenty patterns composed of *stimuli of both colours* (for free-choice RT's). Three, four or five forced choices separated every two successive free choices. Colours, positions and combinations of positions of stimuli all appeared equally often in a random order. One sequence, the '*single*' sequence, used *one* spot of light for each forced choice and *two* spots, one of each colour, for free choices. The other, the '*double*' sequence, used *two* neighbouring spots of the same colour for forced choices and *four* spots, including two adjacent spots of each colour, for free choices.

Both red and green spots had luminances of approximately 2 millilamberts. The 'double' sequence may be assumed to have aroused stronger response tendencies, whether one regards an increase in extensity as equivalent to an increase in intensity, on the grounds that both mean an addition to the amount of energy impinging on the retina, or one regards the two neighbouring spots as two stimuli evoking the same response, in which case the 'principle of behavioural summation' (Hull, 1943, p. 213; 1951, pp. 64–5) implies that the response strength would exceed that produced by one stimulus alone.

Every S had two sessions, 24 hr. apart, of which one was with the C response unit and one with the I unit, one with the 'single' stimulus sequence and one with the 'double' stimulus sequence. The S's were divided into four groups of five, who had the experimental conditions in the following combinations and orders respectively: C single, I double; I double, C single; C double, I single; I single, C double.

Procedure. When the I response unit was used, S grasped the switch between the thumb and forefinger of his right hand. He was told to push the switch forward as soon as he saw a green light and backward as soon as he saw a red light. The switch was not to be returned to the neutral position until the light went off. When lights of both colours appeared, either response, but not both, was to be performed. With the C response unit, the task was similar, except that both hands were to be used; the right-hand switch was to be pushed forward for a green light and the left-hand switch forward for a red light. A few forced choices were given for practice at the start of each session.

Results. For each session with each S, mean RT's for forced choices with green stimuli, forced choices with red stimuli and free choices were calculated. The data are exhibited in Table 1.

Table 1. *Experiment I. Mean reaction times in milliseconds*

	Forced choices			Free choices
	Green	Red	All	
Overall mean	1034	1002	1018	1270
Between S's s.d.	—	—	*199*	*323*
Double stimuli	1000	989	994	1323
Single stimuli	1069	1015	1042	1216
I response unit	1072	1055	1063	1270
C response unit	992	948	973	1269
First session	1043	975	1009	1265
Second session	1026	1028	1027	1274

The prediction that free-choice RT's would be longer than forced-choice RT's was investigated with Wilcoxon's T-test for paired replicates (Mosteller & Bush, 1954). The free-choice mean was higher than the forced-choice mean for 19 out of 20 S's, so that $T=1$ and $p<0{\cdot}01$.

Split-plot analyses of variance (Cochran & Cox, 1950) were carried out on forced-choice and free-choice means separately. These analyses showed that the 'double' stimulus sequence produced *shorter forced choice* ($F=4{\cdot}77$, d.f. $=1, 17$, $p<0{\cdot}05$), but *longer free-choice* RT's ($F=11{\cdot}92$, d.f. $=1, 17$, $p<0{\cdot}01$) than the 'single' stimulus sequence. The former finding is, no doubt, related to the well-known fact that more intense stimuli produce shorter simple RT's (Hull, 1949), while the latter is in line with our expectation that an increase in the absolute strengths of competing response tendencies will increase conflict. It should be noted that the free-choice RT's for double stimuli are not only greater in absolute value than those for single stimuli. They also represent a greater percentage-wise increase over the corresponding forced-choice RT's, the free-choice RT's being 33·1 % longer than the forced-choice RT's for 'double' stimuli and 16·7 % longer for 'single' stimuli.

A comparison of RT's with the two response units showed that the physiologically incompatible pair of responses (i.e those required by the I unit) produced significantly longer forced-choice RT's ($\mathrm{F}=15{\cdot}36$, d.f. = 1, 17, $p<0{\cdot}01$), but there was no difference in the case of free-choice RT's ($\mathrm{F}=0+$). Since the two different response units necessitated different responses to red but identical responses to green stimuli, an examination of forced-choice RT's to green only constitutes the fairest test of the degree-of-incompatibility variable. This, once again, shows the I response unit to have produced longer forced-choice RT's ($F=10{\cdot}77$, d.f. = 1, 17, $p<0{\cdot}01$).

There were two other interesting results. One was the significant interaction between sessions and colours of stimuli for forced choices ($F=7{\cdot}68$, d.f. = 1, 17, $p<0{\cdot}05$). RT's to red stimuli were shorter in the first session, but they lost this advantage in the second session. This was, no doubt, due to the fact that different responses to red were demanded in the two sessions using different response units. In the second session, therefore, there would be three competing responses evoked by red stimuli: the correct response, the generalized response appropriate to green, and the interfering response learned in the first session. That the RT's increased as a result is a further confirmation of the importance of the number of competing response tendencies variable. Secondly, the mean forced-choice RT for each S and the difference between his mean free-choice RT and his mean forced-choice RT were taken, and the product-moment correlation was computed. This came to $+0{\cdot}63$, $t=3{\cdot}47$, d.f. = 18, $p<0{\cdot}01$. This suggests that an investigation of relations between the two types of choice RT and personality factors may be worth while. A S's forced-choice RT and the extent to which this is exceeded by his free-choice RT may both reflect some characteristic reaction to conflict.

III. EXPERIMENT II

A second experiment was undertaken for two purposes. First, it was felt that the degree-of-incompatibility factor deserved another test. In Exp. I, different response units were used for physiologically compatible and incompatible pairs of responses. This meant that S's might possibly have changed their postures for the two conditions, which might have affected their RT's, although it would be difficult on this hypothesis to explain why free-choice RT's were not influenced in the same way. Secondly, it was thought desirable to see what happened to free-choice RT's when stimuli were fainter than any used in Exp. I.

Apparatus was the same as in Exp. I, apart from the following modifications. (1) The *stimulus unit* was timed to present stimuli for 4 sec. at 4 sec. intervals. White and red spots of light were used. For half of the S's the two left-hand holes and for the remainder the two right-hand holes were covered with six sheets of ordinary white typing paper, which made the spots appearing there about as faint as they could be without the colours becoming hard to distinguish. The spots appearing at the two uncovered holes remained at the order of intensity that was used for the single stimulus sequence in Exp. I. The approximate luminances in millilamberts were 6 for the bright white spots, 2 for the bright red spots, 0·20 for the dim white spots, and 0·06 for the dim red spots. (2) One *response unit* was used. It consisted of two switches mounted on a board, both of which could be moved backward or forward.

Subjects were ten boys and six girls, all aged 10 years and attending a primary school.

Design and Procedure. Two stimulus sequences were drawn up, each composed of sixteen presentations of a *red* spot of light, sixteen of a *white* spot and sixteen presentations of *a red and a white spot together*, randomly interspersed. All four positions for the forced-choice stimuli were used equally often in a random order, as were the right-hand and the left-hand pairs of positions for free choices. Half of the forced choices and half of the free choices were thus with bright stimuli and the remainder with dim stimuli.

Each S had two sessions of forty-eight responses, one with each stimulus sequence, in immediate succession. He was instructed to grasp both switches throughout. One session, the first for half of the S's and the second for the remainder, was with condition I (physiologically incompatible responses), i.e. the right-hand switch was to be moved forward in response to a white light and backward in response to a red light. The other session was with condition C (physiologically compatible responses), i.e. the right-hand switch was to be moved forward in response to a white light and the left-hand switch was to be moved forward in response to a red light. When lights of both colours were presented together, either response was to be performed, as with the free choices of Exp. I.

Results. Mean RT's for the various conditions are shown in Table 2.

Table 2. *Experiment II. Mean reaction times in milliseconds*

	Forced choices			Free choices
	White	Red	All	
Overall mean	1208	1190	1199	1428
Between S's S.D.	—	—	*229*	*326*
Bright stimuli	1099	1110	1105	1377
Dim stimuli	1316	1270	1293	1478
I responses	1232	1240	1235	1440
C responses	1184	1140	1162	1415
First session	1202	1188	1195	1454
Second session	1214	1192	1203	1401

A single analysis of variance was carried out on the *total* reaction times for the various conditions. The free-choice totals were doubled to make them comparable with the forced-choice totals, and they were then significantly higher than the latter at the 0·001 level ($F=22{\cdot}62$, d.f. $=1$, 15). The difference between I responses and C responses was not significant ($F=2{\cdot}60$, d.f. $=1$, 14). But when forced-choice RT's to white stimuli only (which evoked identical responses in both conditions) were analysed separately, F came to 3·53, and $p<0{\cdot}05$ by a one-tailed test. We have therefore obtained some confirmation of the conclusion, supported by Exp. I, that physiological incompatibility between responses lengthens forced-choice RT's. But once again no evidence was found for a similar effect on free-choice RT's. The interaction between I or C responses and forced or free choices is insignificant.

It will be seen from Table 2 that both types of RT were longer with dimmer stimuli, and, in fact, F for the intensity variable came to 47·91, which, with 1 and 42 d.f., exceeds the 0·001 significance level. But the interaction between forced or free choices and bright or dim stimuli produced an F of 4·47, which, with 1 and 42 d.f., is significant at the 0·05 level, showing that the intensity variable had a smaller effect on free-choice RT's. At first sight, the results of Exp. II seem to contradict those of Exp. I, since, in Exp. II, the dimmer stimuli produced the longer free-choice RT's. But if we examine the percentage by which free-choice RT's exceeded forced-choice RT's we find the same trend as before. In Exp. I the percentage was 33·1 for the more intense stimuli and 16·7 for the less intense. The brighter stimuli of Exp. II (which approximate the less intense stimuli of Exp. I) produce a percentage of 24·6 (the difference being explicable by the change in S's and conditions), and the dimmer stimuli a percentage of 14·3.

It seems likely that slower eye movements were at least partly responsible for the lengthening of both types of choice RT with fainter stimuli. Receptor orientations and possibly other sensory or neural processes necessary for dealing with a stimulus may take a certain minimum time, so that only part of a choice RT may reflect the degree of conflict between the switch-pressing responses. The relative slowness of eye movements, etc., with fainter stimuli may itself be attributable to conflict. Receptor orientations can be regarded as forced-choice reactions, and, the fainter the spots of light, the more nearly equal the tendencies to fixate the four apertures.

The correlation between forced-choice RT's and the excess of free-choice over forced-choice RT's for individual S's came to +0·24, which is not significant. But the analysis of variance showed a significant interaction between S's and forced or free choices ($F = 5{\cdot}31$, d.f. $= 15, 42$, $p < 0{\cdot}01$).

IV. EXPERIMENT III

Exp. III also had two purposes. The first was to verify that an increase in the number of alternative responses would lengthen free-choice RT's as well as forced-choice RT's. The second was to find out whether Eysenck's neuroticism (P) factor (1947) would affect performance. It was thought that individuals with higher neuroticism might show a higher proportional increase of free-choice over forced-choice RT's.

Apparatus was as in the two previous experiments, except that only green spots appeared at the two end positions and only red spots at the two middle positions. The same luminances as in Exp. I were used, and a vertical line of luminous paint divided the two right-hand from the two left-hand apertures. The room was almost completely dark. The *response unit* was that used in Exp. II.

Subjects were twenty-four undergraduates, selected as follows: The Maudsley Personality Inventory was given to thirty-four men and forty women attending a first-year psychology lecture. The scores on the Neuroticism Scale and the Lie Scale were computed. Eight high- and low-neuroticism S's who had recently participated in another experiment were omitted, as were two low-neuroticism men with a score of more than 10 out of 20 on the Lie Scale and one low-neuroticism man who failed to respond to a request to take part. Apart from these, the twenty-four S's selected consisted of the six men and the six women with the highest and the six men and the six women with the lowest neuroticism scores. The ranges for these four groups were (out of 24) 16·5–21·5, 18·5–23, 2·5–6·5 and 7–9 respectively.

Design and procedure. Every S went through a *four-alternative* sequence, in which all four apertures of the stimulus unit were used, and a *two-alternative* sequence, in which only the two apertures on one side of the central dividing line were used. Three S's in each group had the two-alternative sequence, with the two right-hand positions, before the four-alternative sequence. The remaining three had the four-alternative sequence first, and in their case the two-alternative sequence used the two left-hand positions.

Both sequences comprised thirty-six stimulus patterns. The two-alternative sequence included twelve presentations of a *green* spot in one position, twelve of a *red* spot in the other position and twelve of *both together*. The four-alternative sequence included twelve presentations of a *green* spot, the two end positions being used equally often, twelve of a *red* spot, the two middle positions being used equally often, and twelve of *two green and two red spots simultaneously*. The first four patterns of the four-alternative sequence consisted of one presentation of each single stimulus, but otherwise the sequences were randomized.

S was instructed to respond with the key that was on the same side as the spot of light, so that only one key was used when there were two alternatives, but both keys when there were four alternatives. The key was to be pressed forward in response to a green stimulus and backward in response to a red one. There were thus as many alternative responses as there were alternative stimuli. When two or four spots lit up together, S was to respond to any one of the stimuli and ignore the rest.

Results. The mean forced-choice and free-choice RT's in both two-alternative and four-alternative conditions were calculated for each S. These means are displayed in Table 3. They were subjected to a logarithmic transformation before undergoing analysis of variance, because the standard deviations increased steeply with the means.

The analysis showed that free-choice RT's were significantly longer than forced-choice RT's ($F=22{\cdot}64$, d.f. $=1$, 48, $p<0{\cdot}001$) and four-alternative RT's were significantly longer than two-alternative RT's ($F=276{\cdot}78$, d.f. $=1$, 48, $p<0{\cdot}001$). The interaction between these two variables did not approach significance ($F<1$). We can therefore conclude that both forced- and free-choice RT's increase when the number of alternative responses rises but, in view of the logarithmic transformation, that the proportion by which the latter exceed the former is not significantly different whether there are two or four alternatives. This proportion was actually 10·1% with two and 10·4% with four alternatives.

None of the comparisons or interactions involving neuroticism or sex reached significance, with the sole exception of the neuroticism × order interaction ($F=10{\cdot}32$, d.f. $=1$, 16, $p<0{\cdot}01$). The two-alternative to four-alternative sequence produced lower RT's than the other sequence in the case of the high-neuroticism S's but higher RT's in the case of the low-neuroticism S's.

Table 3. *Experiment III. Mean reaction time in milliseconds*

	Two alternatives		Four alternatives	
	Forced choices	Free choices	Forced choices	Free choices
All S's	585	644	807	891
Between S's S.D.	*97*	*141*	*183*	*207*
High-neuroticism S's	589	633	836	886
Low-neuroticism S's	581	654	779	896

V. DISCUSSION

The theoretical considerations with which this paper began have performed one of the principal functions of theory: they have directed our attention to variables whose importance was subsequently confirmed by experimental data. It now remains to discuss the significance of the findings further.

They are in general accord with the conclusion that choice time, in so far as it exceeds the minimum time for registration of a stimulus, increases with three of the proposed determinants of degree of conflict, viz. the absolute strength, the nearness to equality and the number of competing response tendencies. The physiological compatibility-of-responses variable, however, turned out to affect forced-choice but not free-choice RT's. The most likely explanation for this last fact seems to be as follows. Let us suppose that a S is in a forced-choice situation and begins to make the wrong response. If the correct response consists of moving the same hand in a different direction, it cannot be initiated before the wrong response has been cancelled. But if the correct response consists of a movement with the other hand, it can be initiated earlier. The quicker correction of errors may well explain the shorter average RT's with the latter condition. With free-choice RT's, on the other hand, either response is acceptable, and so the question of correction does not arise.

At all events, it seems doubtful whether the degree of *physiological* incompatibility should continue to be included among the determinants of degree of conflict. There seem, however, to be cases of *acquired* incompatibility, e.g. between symbolic responses participating in thought processes (Berlyne, 1954*b*), where the extent of the incompatibility is of importance. Miller (1944, p 458) describes two relevant types of acquired conflict. In the first type, 'the individual learns to respond with anxiety to the pattern produced by simultaneous stimulation from both responses'. The degree of incompatibility can, in such instances, be identified with the strength of the anxiety-mediating habit. In the second type, 'the subject learns to react directly to the stimuli produced by each of (the responses) with a response incompatible with the other', i.e. if the two original responses are R_1 and R_2, the proprioceptive and other stimuli resulting from R_1 come to evoke a response R_{-2}, which is incompatible with R_2, and stimuli resulting from R_2 come to evoke R_{-1}, which is incompatible with R_1. When this happens, the degree of incompatibility can be identified with the ratio of the strength of R_{-1} to that of R_1 and the ratio of the strength R_{-2} to that of R_2 and thus reduced to the equality-of-strength variable.

We can now turn to various theoretical models, treating phenomena closely akin to those that are our concern here, and see whether they can account for the relations between the determinants of degree of conflict and reaction time. It may be enlightening in this connexion to consider not why free-choice RT's are relatively long but why S's ever come to react at all. What must occur before one of the competing reactions can prevail over the other? The different models suggest different answers.

(1) One possible answer is that some relatively improbable combination of events must come about before the conflict can be resolved, and that at least part of the mean RT consists of the time that must elapse before chance can be expected to produce this combination.

According to Hull's principle of behavioural oscillation (1943), response tendencies have 'momentary effective reaction potentials ($_S\dot{\bar{E}}_R$)', which fluctuate independently with approximately normal distributions. A response will not be performed unless its $_S\dot{\bar{E}}_R$ is higher than the 'reaction threshold'. There are three models that can be based on these assumptions.

(*a*) Hull's own statement about competing response tendencies appears as Postulate 16 in *Principles of Behaviour*. It is that only the response with the highest $_S\dot{\bar{E}}_R$ will occur. But when it occurs, there is no reason why it should not have the same latency as if it had been evoked alone, since latency (Postulate 13) depends only on $_S\dot{\bar{E}}_R$ (or according to later versions of the theory (1951, 1952) on $_SE_R$). Moreover, it should take no longer for one of two strong response tendencies than for one of two weak ones to have a higher $_S\dot{\bar{E}}_R$.

(*b*) An alternative model is provided by the plausible additional assumption that an overt response will not occur until the fluctuations have brought one response tendency above the reaction threshold at a time when the other response tendencies are below it. Let us assume that the probability of the $_S\dot{\bar{E}}_R$ of a certain response exceeding the threshold in a particular unit of time is p. The portion of the mean simple RT which depends on behavioural oscillation would then be $1/p$ units. Let us now suppose that we have two such responses competing. The probability that either will be above-threshold while the other is not will be $2p(1-p)$, and the corresponding portion of the mean free-choice RT will be the reciprocal of this expression. Now the value of $\dfrac{1}{2p(1-p)}$ reaches a minimum when $p=0{\cdot}5$. The free-choice RT will be longer with stronger response tendencies, because both tendencies will be above the threshold for most of the time, and it will be longer with weaker response tendencies, because both tendencies will be below the threshold for most of the time. The curvilinear relationship between free-choice RT and stimulus intensity that is suggested by Exps. I and II fits this implication of the model very well, but unfortunately there is another implication that does not agree with the data. This is that, when $p<0{\cdot}5$, $\frac{1}{p}>\frac{1}{2p(1-p)}$, which means that the simple RT will be longer than the free-choice RT. Since forced-choice RT's are known to exceed simple RT's, it follows *a fortiori* that forced-choice RT's will exceed free-choice RT's when response tendencies are weak enough to bring the latter above their minimum. This conclusion is not borne out by the results of Exp. II, and so the model is not tenable. (This model might possibly be applicable to situations in which there are very weak habits. For example, 'name the capital of Great Britain' might conceivably produce a quicker response than 'name any city beginning with L', whereas 'name the capital of Paraguay' might produce a slower response than 'name any South American capital'.)

(*c*) Another possibility is that the ${}_S\dot{\bar{E}}_R$ of one response must exceed those of all competing responses by a certain minimum quantity k before overt action can take place. The probability of this happening will be least, and consequently the average number of time units before it happens will be greatest, when the response tendencies are equal in strength. The probability will also diminish, and the reaction time will therefore go up, when the number of response tendencies is increased. In order to predict that the part of the free-choice RT dependent on behavioural oscillation will increase with the absolute values of the ${}_S\bar{E}_R$'s, we have to add the postulate, analogous to the Weber–Fechner law, that k is larger with stronger response tendencies. If the intensity of one stimulus has to exceed that of another by a certain proportion before an organism reacts to the difference, it is not unreasonable to suppose that random fluctuations must make the ${}_S\dot{\bar{E}}_R$ of one response exceed those of its competitors by a certain proportion before it can eclipse them. The quantity k would thus bear the same relation to the differential sensory threshold as Hull's reaction threshold (${}_SL_R$) bears to the absolute sensory threshold. The fluctuations that give one of two equal response tendencies this temporary advantage would be analogous to those that cause one of two equal stimuli to be judged more intense than the other in a psychophysical experiment. This model seems, so far, to be the most promising of those derived from the concept of behavioural oscillation.

All these probabilistic models imply some variation in the responses a S will select in successive exposures to a free-choice situation. Various factors may make one of the alternative responses more frequent than another, but the models could scarcely stand up if a S always made the same response. In order to verify that this does not happen, a supplementary experiment was carried out. The S's were six boys from the same school and age group as the S's for Exp. I. All of them received the '*single*' sequence from Exp. I with 3 sec. on to 3 sec. off timing. Half performed the C responses and the other half the I responses. The mean number of responses to red stimuli, in the red-green free-choice situations, was 8·5 out of 20, the range being 3–12.

(2) According to Miller's (1944) classification, our free-choice situation produces an approach-approach conflict. A choice should be made in favour of one alternative as soon as random variation of behaviour has inclined the organism towards one of the

alternatives. Miller's theory does not imply that this would take longer to occur when stronger approach tendencies are competing. An assumption might, however, be added to the effect that a larger inclination in one direction, which would presumably take a longer time to occur, is needed to tip the scales when the conflicting approach gradients are higher. This would produce a model closely related to model 1 (*c*).

(3) According to Hick's information-theory model for the forced-choice situation (1952, 1954), the organism absorbs information at a more or less constant rate and is ready to act when its uncertainty has fallen to a certain minimum, comparable to a statistical significance level. This model does not seem to be applicable to the free-choice situation. In particular, it does not explain why free-choice RT's exceed forced-choice RT's or why the former vary with the absolute strength of response tendencies.

The resolution of free-choice conflicts, as Platt (1956) points out, means a loss of information both by 'equivocation' (since several combinations of stimuli can result in the same response) and by 'noise' (since one stimulus pattern can give rise to any of several responses) (Shannon & Weaver, 1949, chap. II). The S has to discard some information, as we are all constantly doing when we concentrate our attention on one part of the environment and make no use of the rest of the information that is being showered on our sensory surfaces. He also has to make a decision which creates information. These processes appear to take time.

(4) A model of yet another sort is suggested by Hebb's (1949) theory of emotion. The 'disruption of cortical organization' which underlies 'emotional disturbance' may, Hebb argues, result from conflict between incompatible phase sequences. The disruption 'may vary in extent *and* in duration', leading 'when it has more than momentary duration', to 'some inco-ordination of effectors', as Hebb's studies of rage and fear, as well as the work of Desai (1939) and Luria (1932), illustrate. The hypothesis might be ventured that the duration of cortical disruption increases with the variables we have cited as determinants of degree of conflict and that responses are delayed until disruption has subsided.

This model derived from Hebb's theory may well be capable of integration with one of the others. It needs, in fact, to be supplemented, if only to explain how one response comes to be chosen when action is no longer blocked. Emotional and motivational factors may not be out of place, for example, in information-theory analyses. Hick (1952) found that forced-choice RT's went down when S's were allowed to make some errors. If S's refrain from responding until their uncertainty with respect to the identity of the stimuli is low enough, the amount of uncertainty they are prepared to tolerate may plausibly depend on how much they are afraid of making a wrong decision. Similarly, our S's had markedly higher RT's, for both forced and free choices, at the beginnings of sequences than later, and we may surmise that a higher level of anxiety on approaching an unfamiliar task was responsible. On the other hand, we failed to find any difference between S's at opposite ends of the distribution for 'neuroticism', a trait associated with proneness to emotional disturbance.

Cerebral processes are thought by Hebb to be so susceptible to disruption because each depends on delicate synchronization between two or more antecedent processes. The probabilistic models discussed under (1) likewise make overt action depend on an appropriate relation obtaining between two or more variables at a particular moment. It may thus be possible to merge model 4 with model 1 (*c*), by postulating that cortical disruption will last until the superiority of one ${}_S\dot{\bar{E}}_R$ over its rivals reaches the 'differential reaction-threshold'.

The author would like to acknowledge the generous co-operation of the Headmasters of Old Aberdeen Junior Secondary School and Aberdeen Demonstration School, who arranged for subjects and rooms, and of Mr C. J. Stuart, the Psychology Department technician, who designed and constructed the apparatus.

REFERENCES

BAKER, L. M. & ELLIOTT, D. N. (1948). Controlled and free association times with identical stimulus- and response-words. *Amer. J. Psychol.* **61**, 535–9.

BARKER, R. G. (1942). An experimental study of the resolution of conflict by children: time elapsing and vicarious trial-and-error behavior occurring. (In McNemar, G., and Merrill, M.A., *Studies in Personality*. New York: McGraw-Hill.)

BERLYNE, D. E. (1950). Stimulus intensity and attention in relation to learning theory. *Quart. J. Exp. Psychol.* **2**, 71–5.

BERLYNE, D. E. (1951*a*). Attention, perception and behaviour theory. *Psychol. Rev.* **58**, 137–46.

BERLYNE, D. E. (1951*b*). Attention to change. *Brit. J. Psychol.* **42**, 269–78.

BERLYNE, D. E. (1954*a*). A theory of human curiosity. *Brit. J. Psychol.* **45**, 180–91.

BERLYNE, D. E. (1954*b*). An experimental study of human curiosity. *Brit. J. Psychol.* **45**, 256–65.

BERLYNE, D. E. (1957). Attention to change, conditioned inhibition (${}_SI_R$) and stimulus satiation. *Brit. J. Psychol.* **48**, 138–40.

BROWN, J. S. & FARBER, I. E. (1951). Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.* **48**, 465–95.

BRUNER, J. S. & POSTMAN, L. (1949). On the perception of incongruity: a paradigm. *J. Pers.* **18**, 206–23.

COCHRAN, W. G. & COX, G. M. (1950). *Experimental Designs*. New York: Wiley.

CROSSMAN, E. R. F. W. (1953). Entropy and choice time: the effect of frequency unbalance on choice-response. *Quart. J. Exp. Psychol.* **5**, 41–51.

DEMBER, W. N. (1954). *Decision Time as a Distance Function*. University of Michigan (Duplicated).

DESAI, M. M. (1939). Surprise. *Brit. J. Psychology. Monogr. Suppl.* **22**.

EYSENCK, H. J. (1947). *Dimensions of Personality*. London: Kegan Paul.

FREUD, S. (1929). *Introductory Lectures on Psycho-analysis*. London: Allen and Unwin.

GIBSON, E. J. (1940). A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.* **47**, 196–229.

HEBB, D. O. (1949). *The Organization of Behavior*. New York: Wiley.

HICK, W. E. (1952). On the rate of gain of information. *Quart. J. Exp. Psychol.* **4**, 11–26.

HICK, W. E. (1954). The impact of information theory on psychology. *Advanc. Sci.* **40**, 397–402.

HULL, C. L. (1943). *Principles of Behavior*. New York: Appleton-Century.

HULL, C. L. (1949). Stimulus-intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.* **56**, 67–76.

HULL, C. L. (1951). *Essentials of Behavior*. New Haven: Yale University Press.

HULL, C. L. (1952). *A Behavior System*. New Haven: Yale University Press.

HYMAN, R. (1953). Stimulus information as a determinant of reaction-time. *J. Exp. Psychol.* **45**, 188–96.

JUNG, C. G. (1918). *Studies in Word Association*. London: Heinemann.

KLEMMER, E. T. & MULLER, P. F. (1953). The rate of handling information. *HFORL Memo. Rep.* no. 34.

KONORSKI, J. (1948). *Conditioned Reflexes and Neuron Organization*. Cambridge University Press.

LOGAN, F. A. (1954). A note on stimulus-intensity dynamism (*V*). *Psychol. Rev.* **61**, 77–80.

LURIA, A. R. (1932). *The Nature of Human Conflicts*. New York: Liveright.

MILLER, N. E. (1944). Experimental studies of conflict. (In Hunt, J. McV. (ed.), *Personality and the Behavior Disorders*. New York: Ronald.)

MOSTELLER, F. & BUSH, R. R. (1954). Selected quantitative techniques. (In Lindzey, G. (ed.), *Handbook of Social Psychology*. Cambridge, Mass.: Addison-Wesley.)

PAVLOV, I. P. (1927). *Conditioned Reflexes*. Oxford University Press.

Platt, J. R. (1956). Amplification aspects of biological response and mental activity. *Amer. Sci.* **44**, 180–97.

Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Shipley, W. C., Coffin, J. I. & Hadsell, K. C. (1945). Affective distance and other factors determining reaction time in judgments of color preferences. *J. Exp. Psychol.* **35**, 206–15.

Wilcocks, R. W. (1928). The effect of an unexpected heterogeneity on attention. *J. Gen. Psychol.* **1**, 266–319.

Woodworth, R. S. (1938). *Experimental Psychology*. New York: Holt.

(*Manuscript received* 16 *January* 1956)

PERSONALITY FACTORS AND THE RATE OF CONDITIONING

By CYRIL M. FRANKS

University of London, Institute of Psychiatry, Maudsley Hospital

Sixty normal male students were conditioned, using the eyeblink reflex. It was found (1) that those students who were introverted conditioned considerably better than those who were extraverted, and (2) that there was no correlation between conditioning and neuroticism. This experiment was designed primarily to provide support for an introversion-extraversion theory of conditioning as opposed to Spence's drive theory. The implications of the experiment are discussed and related to other experiments designed to show that Hull's Seventh Postulate, as used by Spence, leads to erroneous deductions. It would seem that conditioning is independent of drive level where the drive is a secondary factor in the conditioning situation.

I. INTRODUCTION

Hull (1943, p. 240 ff.) offers evidence to suggest that behaviour is a function of two principal classes of variable—learning and motivational. These two variables are supposed to combine multiplicatively according to the formula, response$=f({}_SE_R)=f(D\times{}_SH_R)$, where ${}_SE_R$, D and ${}_SH_R$ represent three theoretical constructs, namely, reaction potential, drive strength and habit strength (Postulate no. 7). According to Hull ${}_SH_R$ is a dependent hypothetical learning variable, and its value can be calculated (Hull, 1943, p. 178 ff.); in practice this would seem to be impossible since it is neither clear what his four critical constructs (k, j, u and i) are, nor how they should be determined. In addition, Hull states that there are some other independent variables which may affect the value of ${}_SH_R$ and which cannot at present be defined. D is also a hypothetical construct and is supposed to represent the total drive strength operating in the organism at a given moment, thus in any specific situation the value of D is assumed to be determined by both relevant needs and the total aggregate strength of existing irrelevant needs.

Spence and his colleagues (e.g. Spence & Farber, 1953; Spence & Taylor, 1951; Taylor, 1951) have adapted Hull's model, in particular the above equation, to make predictions about eyelid conditioning. They argue that the total effective drive strength (D) is in part a function of the level of internal anxiety or emotionality of the subject; consequently subjects with a greater degree of anxiety would possess more drive and therefore form conditioned responses better than subjects with a lesser degree of anxiety. They have used Taylor's Manifest Anxiety Scale (Taylor, 1953) as their measure of anxiety, and claim that, in general, this prediction has been confirmed. (This scale and its relationship to conditioning is discussed elsewhere in this paper.)

An entirely different theory has been developed by Eysenck (1955*a*, *b*) on the basis of certain Pavlovian concepts and hypotheses. (Pavlov's original observation (1927, pp. 397 f.) was to the effect that neurasthenics possessed an exaggeration of the excitatory process and hysterics an exaggeration of the inhibitory process.) Retaining the notion of excitation and inhibition potentials as determining ${}_SE_R$, Eysenck has postulated individual differences in the development of these two potentials and linked these differences with personality development according to the following postulate: 'Individuals in whom reactive inhibition is generated quickly, in whom strong reactive inhibitions are generated, and in whom reactive inhibition is dissipated slowly are thereby predisposed to develop extraverted patterns of behaviour and to develop hysterico-psychopathic disorders in cases of neurotic breakdown; conversely, individuals in whom reactive inhibition has developed slowly, in whom weak reactive inhibitions are generated and in whom reactive inhibition is dissipated quickly, are thereby predisposed to develop introverted patterns of behaviour and to develop dysthymic disorders in cases of neurotic breakdown' (Eysenck, 1955*a*, p. 35).

The well-established finding (e.g. Spence and his colleagues; Welch and his colleagues, 1947*a*, *b*) that anxious subjects condition better than non-anxious may be explained either in terms of Spence's theory of superior drive level or in terms of Eysenck's theory of inhibition-excitation and extraversion-introversion. (Eysenck (1952, 1953) has established that anxious subjects are more introverted and more neurotic than normal subjects.) A crucial experiment would be one for which each theory would predict different results. One such experiment is that in which drive is manipulated, while introversion-extraversion remains unaltered; Spence's theory would predict a change in conditionability, Eysenck's theory would predict no change. To increase drive a group of normal male subjects were deprived of food, drink and tobacco for 24 hr., while another similar group were allowed to eat, drink and smoke in their usual manner. It was found that both groups conditioned equally well (Franks, 1957).

Another crucial experiment would be one which examined the conditionability of three groups, namely, normals, introverted neurotics (anxiety states), and extraverted neurotics (hysterics and psychopaths). Spence's theory would presumably predict that both groups of neurotics would condition better than the normal group (the assumption being that high neuroticism, or high emotionality, may be associated with high drive). Eysenck's theory would predict that the introverted neurotics would condition best, the extraverted neurotic worst and that the normals—being as a group neither introverted nor extraverted—would be intermediate in their ease and strength of conditioning. Such an experiment was carried out by the present writer (Franks, 1956*a*). It was found, using both the eyeblink reflex and the PGR, that patients diagnosed as anxiety states conditioned readily, and patients diagnosed as hysterics or psychopaths conditioned poorly, the normals being in between. Several other findings are of relevance, in particular, that both neurotic groups had high scores on Taylor's anxiety scale, whereas the normal group had a low score on this scale. Furthermore, like other workers (Bitterman & Holtzmann, 1952; Hilgard, Jones & Kaplan, 1951), the writer found no significant correlation between conditioning and Taylor's scale. It was also found that conditionability was unrelated to neuroticism as measured by the Maudsley Medical Questionnaire or by Guilford's *D* and *C* scales (1940), but was closely related to introversion-extraversion, as measured by Guilford's rhathymia scale, *R* (1940). (The evidence for the Maudsley Medical Questionnaire and Guilford's *D* and *C* scales being measures of general neuroticism and Guilford's rhathymia scale being a measure of extraversion is presented by Eysenck (1952, 1953) together with a general discussion of these scales. Guilford's scales are also discussed elsewhere in the present article.) Spence's hypothesis, therefore, appears to gain no support from the experimental findings detailed above, and it seems reasonable to conclude that when Hull's equation, *as used by Spence*, is applied to conditioning situations where the drive is irrelevant, then conditioning is dependent not upon the drive component but only upon the excitation component, i.e. upon ${}_SH_R$.

Two objections to these findings may be raised. First, whereas Spence and his colleagues worked mainly with normal student volunteers, the writer used mainly hospitalized neurotic patients. Secondly, since it was impossible to avoid knowing or guessing into which of the two categories each patient fell it could be possible that attitudinal factors were systematically influencing the results. Therefore it would seem desirable to carry out a similar study on a group of normal student volunteers about whom nothing relevant was known and, as a further precaution, have a different experimenter administer the questionnaires.

II. PLAN OF PRESENT EXPERIMENT

Subjects. Sixty paid undergraduate male volunteers were used ranging in age from 18 to 26 years. All subjects were given a meal allowance and travelling expenses. None of them was studying psychology as a main subject, and all of them were naïve with respect to experimental methods in psychology.

Conditioning apparatus. The conditioning apparatus and laboratory has been described in detail elsewhere (Franks, 1956*a*). Briefly, the unconditioned stimulus was an air puff of 500 msec. duration and pressure 65 mm. of mercury, delivered from a 2·5 mm. internal diameter polythene tube situated approximately 2 cm. from the right eye. The conditioned stimulus was a pure tone, delivered to both ears through a pair of high-quality padded earphones; its frequency was 1100 cycles per sec. and its duration 800 msec., at an intensity of 65 db. above the frequency threshold and so arranged that the air puff began 350 msec. after the start of the tone. All time intervals were accurately controlled by means of a small photoelectric cell (Franks & Withers, 1955) mounted upon the same lens as supported the air-puff delivery tube. The air was supplied from an electronically operated gas cylinder, and the recording milliammeter was so arranged that eyelid movements and the occurrences of the various stimuli all appeared on the same record. The experiment took place in a sound-proof laboratory (Franks, 1955) so arranged that the apparatus and experimenter were in one half of the room and the subjects being conditioned in the other half, the two parts of the room being separated by a sliding curtain.

Personality questionnaire. All subjects were given a new questionnaire, the Maudsley Personality Inventory, recently developed by Eysenck (1957) from the questionnaires of the Guilfords (1934, 1936, 1939*a*, *b*) and from the Maudsley Medical Questionnaire. The need for a new questionnaire arose because the Guilford scales, although they can be used to measure the two orthogonal dimensions of introversion-extraversion and neuroticism, have been observed to possess certain limitations. First, it would seem that in the derivation of these scales little attention has been paid to sex differences; working with these scales suggests that some items in each scale are more predictive of a high score on that scale in the case of men than for women, and vice versa. Secondly, the claim by the Guilfords that each of the scales measures a unitary factor is dubious; it would seem that certain scales are far from unitary and can be resolved into two unrelated components. There is good evidence (Eysenck, 1953) to indicate that when a factor analysis of the intercorrelations between the Guilford scales is carried out two clear-cut orthogonal factors emerge, these two factors being identifiable as neroticism and introversion-extraversion respectively, scales *C* (cycloid disposition) and *D* (depression) most characterising neuroticism, and scales *R* (rhathymia) and—to a lesser extent—*G* (general activity) and *A* (ascendancy) characterizing introversion-extraversion. Other studies strongly support these conclusions (Eysenck, 1955*a*, *b*). Thirdly, although the *R* scale has a low correlation with both the *C* and the *D* scales, an item analysis showed that certain individual items in the *R* scale correlated more highly with the total score on the *C* scale or the *D* scale than with the total score on the *R* scale. A similar deficiency was found in some of the *D* and *C* items.

The new scale was constructed in such a way that these limitations were largely obviated. The neuroticism scale (*N*) and the introversion-extraversion scale (*E*) each contains twenty-four items, so that the test is considerably easier to give and to score than the earlier questionnaires. All items in the *N* scale have significant relations with Guilford's *C* and *D* scales, for both men and women, and insignificant relations with Guilford's *R* scale. All items in the *E* scale have significant relations with Guilford's *R* scale, for both men and women, and insignificant relations with Guilford's *D* and *C* scales (Eysenck, 1956). Sex differences are small for both *E* and *N* scales and the intercorrelation between the two scales is $-0{\cdot}048$ for a sample of 200 men and 200 women. The split half reliabilities of both scales are high.

III. PROCEDURE

The subject was first given the Maudsley Personality Inventory by an independent experimenter. He was then told that records would be made of how well he 'relaxed under various conditions when in a quiet room' and brought to the conditioning laboratory, where he was seated in an armchair, complete with padded head and foot rest. This chair was arranged so that his field of vision was almost completely confined to a small booth immediately in front of him. After the subject was made comfortable, the apparatus was fitted, his hearing threshold taken and the intensity of the tone stimulus adjusted accordingly. Before commencing the conditioning session proper, brief tests were carried out for initial eyeblink sensitivity to sound and for pseudo-conditioning. A technique of partial reinforcement was used, consisting of thirty reinforced trials, interspersed with eighteen test trials (called acquisition trials) consisting of only the conditioned stimulus. After the sequence of thirty reinforcement trials and eighteen acquisition test trials had been given, the subject received a further series of ten successive test trials (called extinction trials). The intertrial interval varied from 20 to 30 sec., and the complete session lasted approximately half an hour.

IV. RESULTS AND DISCUSSION

Of the sixty subjects five had to be rejected for the following reasons: two because of excessive blinking, rendering the records unscorable; one because only irregular unconditioned responses were obtained; one because of an experimental error during testing; and one because the questionnaire was incomplete. For the remaining fifty-five subjects the correlations between conditioning and introversion-extraversion (*E*) were $-0{\cdot}46$ for the number of acquisition conditioned responses and $-0{\cdot}34$ for the number of extinction conditioned responses. These correlations are negative because *E* is a measure of extraversion, and the theory requires the prediction that introverts would condition well and extraverts relatively poorly. The correlations are very similar to those reported between eyeblink conditioning and Guilford's *R* scale of extraversion for a group of sixty subjects consisting mainly of hospitalized neurotics, namely, −0·48 for the acquisition scores and −0·37 for the extinction scores (Franks, 1956*a*). The correlations between conditioning and neuroticism, as measured by the N scale of the Maudsley Personality Inventory, were 0·04 for the acquisition measure and 0·15 for the extinction measure. Using a one-tailed test of significance, the two correlations between conditioning and extraversion are highly significant ($p<0{\cdot}01$), whereas the two correlations between conditioning and neuroticism are insignificant. These results strongly suggest that in normal subjects also conditionability is related to introversion-extraversion and not to neuroticism.

Table 1. *Conditioning and personality scores*

Group		Conditioning		Extraversion (E)	Neuroticism (N)
		Acquisition	Extinction		
All 55 subjects:	Mean	7·18	2·57	13·09	8·90
	S.D.	5·12	2·65	4·49	5·65
High extraversion (15 most extraverted):	Mean	3·73	0·60	1840*·	8·27
	S.D.	3·64	1·14		5·57
Low extraversion (15 least extraverted):	Mean	9·87	3·53	7·53†	9·13
	S.D.	4·87	2·75		4·06
High neuroticism (15 most neurotic):	Mean	5·73	2·27	14·47	16·27‡
	S.D.	5·15	2·85	3·86	
Low Neuroticism (15 least neurotic):	Mean	6·00	2·20	13·60	2·60§
	S.D.	4·02	2·16	3·89	

* This is the mean extraversion score of the top fifteen subjects on the extraversion scale, the range of their scores being from 17 to 22.

† This is the mean extraversion score of the bottom fifteen subjects on the extraversion scale, the range of their scores being from 10 to 3.

‡ This is the mean neuroticism scale of the top fifteen subjects on the neuroticism scale, the range of their scores being from 12 to 22.

§ This is the mean neuroticism score of the bottom fifteen subjects on the neuroticism scale, the range of their score being from 5 to 0.

Table 1 presents the mean number of acquisition conditioned responses given by the fifty-five subjects, together with the mean number of extinction conditioned responses. The mean extraversion score (E) and the mean neuroticism score (N) are also given. Similar data are provided for the fifteen subjects scoring highest on the extraversion scale and the fifteen subjects scoring lowest on this scale, and for the fifteen subjects scoring highest on the neuroticism scale and the fifteen subjects scoring lowest on this scale. Applying t-tests to these data produces the following results:

High versus low extraversion. The high extraversion group give significantly fewer acquisition conditioned responses than the low extraversion group ($t=3{\cdot}79$; $p<0{\cdot}0005$). Similar results were obtained for the extinction conditioning scores ($t=3{\cdot}71$, $p<0{\cdot}0005$). The high and low extraversion groups are very similar in their neuroticism scores and certainly are not significantly different in this respect ($t=0{\cdot}47$).

High versus low neuroticism. The high and low neuroticism groups are very similar in their ease of conditioning and certainly are not significantly different (acquisition, $t=0{\cdot}155$; extinction, $t=0{\cdot}073$). These two groups are also very similar in their extraversion scores ($t=0{\cdot}59$).

Figs. 1 and 2 present the data of Table 1 in a different manner. In Fig. 1, the total number of conditioned eyeblink responses given by the most extraverted fifteen subjects and by the least extraverted fifteen subjects is plotted for each of the eighteen acquisition test trials and each of the ten extinction test trials. It is clear from this figure that the fifteen subjects with the lowest extraversion scores condition much better than the fifteen subjects with the highest extraversion scores. (It might be argued, from Fig. 1, that the more introverted group condition better because they start off better. However, by test trial no. 7 both groups condition equally readily; from then onwards the more introverted group condition far better and furthermore are not completely extinguished after ten successive extinction trials, as are the more extraverted group.) In Fig. 2, the high and low neuroticism groups are compared with respect to their conditionability. It is clear that the fifteen most neurotic subjects closely resemble the fifteen least neurotic subjects in their conditioned response behaviour.

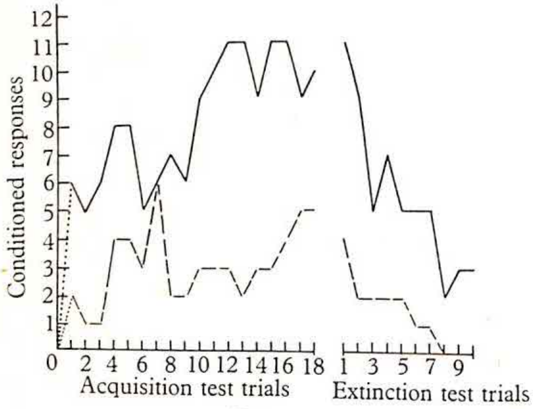


Fig. 1

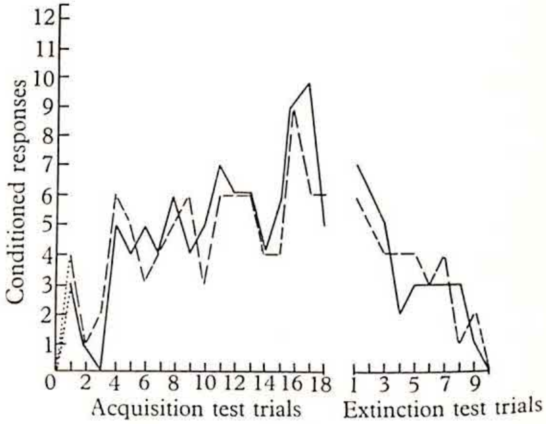


Fig. 2

Fig. 1. The total number of conditioned eyeblink responses given by the fifteen most extraverted subjects (broken line) and by the fifteen least extraverted subjects (solid line).

Fig. 2. The total number of conditioned responses given by the fifteen most neurotic subjects (broken line) and by the fifteen least neurotic subjects (solid line).

From these results it would seem reasonable to conclude that conditionability is closely related to introversion-extraversion, and not at all related to neuroticism, as these two personality dimensions are measured here. Two conclusions would seem to follow from the experiments discussed so far. First, that drive does not affect conditioning in situations where the drive is an irrelevant one. Secondly, that conditioning is unrelated to neuroticism, but closely related to the personality dimension of introversion-extraversion.

The first conclusion requires some elaboration. It is, for example, clear that drives may increase the learning of instrumental responses (e.g. see Munn, 1950, pp. 305 ff.), but it would seem difficult to conceive of eyelid conditioning in the present series of experiments as primarily and consistently instrumental. It is also clear that under certain circumstances, drives increase conditioning when the drives are directly relevant to the conditioning stimuli, thus Zener & McCurdy (1939) found that hungry dogs yield a greater conditioned salivary response than non-hungry dogs, and Lashley (1916) and Winsor (1928) obtained similar results with humans.

The second conclusion is, in part, related to the first, since neuroticism is often regarded as a state of heightened emotionality, and it is not unreasonable to regard emotionality as a form of drive. The second conclusion is of interest also because many workers—quite independent of any drive theory—have expected neurotics to condition better than normals (e.g. Finesinger, Sutherland & McGuire, 1942; Fink, 1943; Schilder, 1929). Further evidence supporting the relationship between conditioning and introversion-extraversion is the finding that sodium amytal, a cortical depressant, reduces conditionability and at the same time increases the degree of extraversion (Franks & Laverty, 1955). This finding is in accord with the general formulation that Pavlov's cortical processes of excitation and inhibition underlie the personality dimension of introversion-extraversion.

If conditioning is closely related to introversion, then it would be expected that a clear positive relationship between conditionability and anxiety would emerge, since anxious subjects have been shown to be introverted. However, Spence and his colleagues obtained only poor relationships between anxiety and eyelid conditioning—their correlations, although positive, were sometimes insignificant and always slight. The explanation would seem to be (Franks, 1956*b*) that their measure of anxiety, the Taylor Anxiety Scale, has an extremely high correlation with the Maudsley Medical Questionnaire, which has been found (Eysenck, 1953) to be an adequate measure of general neuroticism. At the same time, Taylor's scale has a considerably smaller correlation with introversion, this correlation being sufficient to account for the positive but poor results obtained by Spence and his associates. It would seem, then, from the above conclusions and the evidence presented, that Spence's use of Hull's Seventh Postulate leads to erroneous deductions and is, in fact, incorrect in situations where the drive concerned is a secondary factor in the condition situation; what appear to be the determining factors are individual differences in the cortical processes of excitation and inhibition.

REFERENCES

Bitterman, M. E. & Holtzman, W. H. (1952). Conditioning and extinction of the galvanic spin response as a function of anxiety. *J. Abnorm. Soc. Psychol.* **47**, 615–23.

Eysenck, H. J. (1952). *The Scientific Study of Personality.* London: Routledge and Kegan Paul.

Eysenck, H. J. (1953). *The Structure of Human Personality.* London: Methuen.

Eysenck, H. J. (1955*a*) A dynamic theory of anxiety and hysteria. *J. Ment. Sci.* **101**, 28–51.

Eysenck, H. J. (1955*b*). Cortical inhibition, figural after-effect and the theory of personality. *J. Abnorm. Soc. Psychol.* **51**, 94–106.

Eysenck, H. J. (1957). The questionnaire measurement of neuroticism and extraversion (in the Press).

Finesinger, J. E., Sutherland, G. F. & McGuire, F. F. (1942). The positive conditioned salivary reflex in psychoneurotic patients. *Amer. J. Psychiat.* **99**, 61–74.

Fink, D. H. (1943). *Release from Nervous Tension.* New York: Simon and Shuster.

Franks, C. M. (1955). The establishment of a conditioning laboratory for the investigation of personality and cortical functioning. *Nature, Lond.* **175**, 984–5.

Franks, C. M. (1956*a*). Conditioning and personality. *J. Abnorm. Soc. Psychol.* **52**, 143–50.

Franks, C. M. (1956*b*). The Taylor scale and the dimensional analysis of anxiety. *Rev. Psychol. Appl.* **6**, 35–44.

Franks, C. M. (1957). The effect of food, drink and tobacco deprivation upon the conditioned eyeblink response in normal subjects. *J. Exp. Psychol.* **53**, 117–20.

Franks, C. M. & Laverty, S. G. (1955). Sodium amytal and eyelid conditioning. *J. Ment. Sci.* **101**, 654–63.

Franks, C. M. & Withers, W. C. R. (1955). Photoelectric recording of eyelid movements. *Amer. J. Psychol.* **68**, 467–71.

Guilford, J. P. (1940). *An Inventory of Factors STDCR.* Stanford, California: Sheridan Supply Co.

Guilford, J. P. & Guilford, R. B. (1934). An analysis of the factors in a typical test of introversion-extraversion. *J. Abnorm. Soc. Psychol.* **28**, 377–99.

Guilford, J. P. & Guilford, R. B. (1936). Personality factors *S*, *E* and *M*, and their measurement. *J. Psychol.* **2**, 109–27.

GUILFORD, J. P. & GUILFORD, R. B. (1939*a*). Personality factors *D, R, T* and *A*. *J. Abnorm. Soc. Psychol.* **34**, 21–36.

GUILFORD, J. P. & GUILFORD, R. B. (1939*b*). Personality factors *N* and *GD*. *J. Abnorm. Soc. Psychol.* **34**, 239–48.

HILGARD, E. R., JONES, L. V. & KAPLAN, S. J. (1951). Conditioned discrimination as related to anxiety. *J. Exp. Psychol.* **42**, 94–100.

HULL, C. L. (1943). *Principles of Behavior*. New York: D. Appleton-Century.

LASHLEY, K. (1916). Reflex secretion of the human parotid gland. *J. Exp. Psychol.* **1**, 461–93.

MUNN, L. M. (1950). *Handbook of Psychological Research on the Rat*. Cambridge, Mass.: Houghton-Mifflin.

PAVLOV, I. P. (1927). *Conditioned Reflexes* (trans. by G. V. Anrep). London: Oxford University Press.

SCHILDER, P. (1929). The somatic basis of the neurosis. *J. Nerv. Ment. Dis.* **70**, 502–19.

SPENCE, K. W. & FARBER, I. E. (1953). Conditioning and extinction as a function of anxiety. *J. Exp. Psychol.* **45**, 116–19.

SPENCE, K. W. & TAYLOR, J. A. (1951) Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. Exp. Psychol.* **42**, 183–8.

TAYLOR, J. A. (1951). The relationship of anxiety to the conditioned eyelid response. *J. Exp. Psychol.* **41**, 81–92.

TAYLOR, J. A. (1953). A personality scale of manifest anxiety. *J. Abnorm. Soc. Psychol.* **48**, 285–90.

WELCH, L. & KUBIS, J. (1947*a*). The effect of anxiety on the conditioning rate and stability of the P.G.R. *J. Psychol.* **23**, 83–91.

WELCH, L. & KUBIS, J. (1947*b*). Conditioned PGR (Psychogalvanic response) in states of pathological anxiety. *J. Nerv. Ment. Dis.* **105**, 372–81.

WINSOR, A. L. (1928). Conditions affecting human parotid secretion. *J. Exp. Psychol.* **11**, 355–63.

ZENER, K. & MCCURDY, H. G. (1939). Analysis of motivational factors in conditioned behaviour. I. The differential effect of changes in hunger upon conditioned, unconditioned and spontaneous salivary secretions. *J. Psychol.* **8**, 321–50.

(*Manuscript received* 30 *January* 1956)

AN EXPERIMENTAL INVESTIGATION OF SEWING-MACHINE SKILL

By W. T. SINGLETON

British Boot, Shoe and Allied Trades Research Association, Kettering

An attempt was made to measure the transfer effects which occur when the receptor or effector modes of an element of sewing skill are changed. The pre-transfer task required the operation of an industrial sewing machine for about twenty years. Under these conditions no transfer was detected. It is concluded that this particular skill is dependent on the utilization of the particular receptor and effector modes with which it was originally acquired. It is suggested that transfer experiments are not likely to be fruitful unless learning is controlled and, in the case of motor skills, adequate attention is paid to the way in which the peripheral system is being utilized.

I. Introduction

A sensori-motor skill such as that developed by sewing-machine operators is dependent on the timing and precision of limb movements. The correct movements must be made with the correct speed, the correct force, in the correct order, and at the correct time. It is usually found, however, that as the cycles of activity follow each other none of the above remain constant as they would if the task was controlled by a machine. Small changes in the texture of the leather, the size and style of the shoe, etc., which alter the order and timing of the movements do not appear to interfere in the least with the steady progression of the activity. The less competent operator is less economical both of time and of movement. That is, she takes longer over every part of the operation and makes more frequent changes of grip and work position. The problem then is: what is the structure of the skill? Or, to put it another way—what has the skilled operator learned? Or, what particular abilities has she acquired which have not been developed in the unskilled operator? The identification of these so-called key elements of the skill is an important preliminary step in any investigation of ways to improve operators' efficiency, to redesign controls, to train operators or to assess the capacity of new labour.

This skill is too complex to be examined as a whole in a controlled situation. The experiments described below are concerned solely with foot movements and the relevant receptor processes. When the treadle is pressed the machine accelerates, the operator sees the work moving under the needle, she hears the machine noise and the receptor organs in her feet and leg muscles initiate signals concerning what has been done. The higher the skill of the operator the less she needs to attend to all these signals. An inexperienced operator does not know exactly what the machine will do and she has not acquired the ability to grade the muscular response accurately. So she must take time to assimilate the knowledge of results which becomes available, and to correct errors. A highly skilled operator who has made similar responses so often in the past is almost certain about what will happen when she makes the particular response so that all the events which follow convey much less 'information' to her and are therefore more easily and more quickly dealt with.

If we accept this as a reasonable assessment of the situation a further problem arises. By what internal mechanism does the operator control her foot responses and what cues does she utilize? If these questions can be answered even approximately then we have a theoretical base from which to consider ways of reducing the skill required, e.g. by providing better cues or changing the response required (i.e. by changing the controls) so that the precision required of the operator is less severe.

It so happens that during a particular sewing skill known as the Derby side operation, certain quite specific responses are required of the operator. On the standard Gibson upper there are two short rows of stitching which strengthen each flap containing eyelets. A third row at right angles to these along the front edge of the quarter fulfils a similar purpose. These rows are all about eight stitches in length. Thus, in sewing one shoe the operator has, among other tasks, that of making six pulses on the treadle each of which must produce eight stitches. For a particular shoe model the fixed number of stitches must be reproduced accurately—if the row is too long the appearance of the shoe is impaired. For this response it seems reasonable to suggest that experienced operators develop a central mechanism which provides a response of a given time length and which is relatively independent of sensory cues. This can be investigated experimentally using a transfer situation. If such a mechanism exists at a central level then it might be relatively independent of the effector mode of response and should be revealed by positive transfer from one mode to another.

II. Apparatus

A display panel was constructed which consists of twenty-four small neon bulbs equally spaced in a horizontal straight line with a separation of 1 in. between adjacent bulbs. The line of bulbs is inset symmetrically in a dull black surface of size 30 × 12 in. These bulbs can be lit in two ways. (*a*) Using a Morse key which can be controlled by the hand or the foot. When the key is pressed the end light on the left-hand side is lit. This is followed by the next light and so on until the key is released. The speed at which successive lights come on can be varied by the experimenter through three stages which give speeds of 2300, 3000 and 3750 lights per minute. When the lights are set to run at the slowest speed and the key is pressed for 0·21 sec., then eight lights will appear on the display. (*b*) Using a sewing machine which can be controlled either by a standard treadle or by a hand lever. As the machine stitches, lights appear on the screen, running from left to right, at the same speed as that at which stitches are being made. The maximum speed of the machine is 2300 stitches per min. A tachograph is attached to the sewing machine, so that when the treadle is pressed a graphical record of the machine speed during the response is obtained.

The experimenter is provided with another Morse key which will zero the light system. The subject's Morse key is connected so that having made one response, that is, having pressed and released the key to produce a certain number of lights, further pressures on the key have no effect until the system has been returned to zero by the experimenter.

III. Subjects and procedure

Twenty-one female subjects were used. Eleven were from the staff of the Research Association. None of these had any experience in using an industrial sewing machine. The second group of ten were machine operators in a shoe factory, all of whom were accustomed to producing the kind of response required in this experimental situation. Their average length of experience on sewing machines was over 20 years.

Each subject was given the same tasks in the same order. A vertical red line was fixed on the display between the 8th and 9th lights. The subject sat comfortably in front of the display (which was roughly at eye level) with one hand on the Morse key. It was explained that when the key was pressed lights would 'run' from left to right across the display. The subject was to press the key for just long enough for the lights to reach the red line, and a few practice trials were given. When the subject made a response which produced eight lights she was told that this was correct. When she made responses which produced either more or less than eight lights she was told that she was wrong and that she must try to produce eight lights at each attempt. When it was apparent that the subject understood the task she was given three experimental trials, each of sixty-four responses. After each response the lights were returned to the zero position by the experimenter. During the first trial the speed was 2300 lights per minute, during the second trial 3750 per minute, during the third trial 2300 per minute.

The subject was then asked to sit at the sewing machine. Again the display was about 5 ft. in front of her. It was explained that the lights would now effectively work from the sewing-machine treadle and would run at the same speed as that at which the machine was stitching. She was asked to do exactly as she had done in the earlier trials and stop the lights at the red line. She was also asked to make the response as quickly as possible and to try to put on the eight lights in one burst rather than by attempting to stitch very slowly. After a few practice attempts at this the subject was given the fourth and final trial of sixty-four responses using the machine.

About 8 months after the completion of the first part of the experiment the procedure was repeated using the Morse key controlled by a foot pedal and the machine controlled by a hand lever. The same sequence of tasks and, as far as possible, the same subjects were used. Because of staff changes it was necessary to introduce two new subjects into the operator group, but the remaining eight and all the non-operator group were the same as those used for the first part of the experiment.

IV. RESULTS

After each response the experimenter noted the number of the right-hand end light. For this purpose, when eight lights were on the number 0 was noted, when seven or nine lights were on -1 or 1 was noted respectively. Similarly for six lights the number -2 would be recorded, for ten lights $+2$, and so on. The sixty-four scores of one trial can then be plotted as a distribution which appears, in most cases, to be normal in shape. From this distribution various measures of the skill can be derived. One measure is the number of correct responses. In most cases the mode and median of the distribution occurred at a score of zero. A second possible measure is the standard deviation of the distribution which indicates the consistency or the efficiency of 'grouping' of the 'shots'. A third possible measure is the mean of the distribution which reveals any tendency to overshoot or undershoot during the trial. The means of these three measures for each of the Morse key trials are shown in Table 1. Standard deviations of the scores are shown in brackets. The distributions of the subject's scores are such that on none of the three scores is there any significant difference between operators and non-operators. It will also be noted that there is no learning at this task and no difference in performance between trials involving the fast and slow light speeds.

Table 1. *Number in sample: ten operators; eleven non-operators*

		1st Trial (slow)	2nd Trial (fast)	3rd Trial (slow)
Number correct	Operators	13·5 (3·0)	13·1 (7·3)	14·2 (5·4)
	Non-operators	16·2 (7·5)	14·5 (3·2)	17·5 (9·4)
Standard deviation	Operators	2·37 (1·01)	2·21 (0·85)	1·74 (0·51)
	Non-operators	1·85 (0·14)	1·89 (0·58)	1·60 (0·54)
Mean	Operators	0·25 (0·79)	0·33 (0·65)	−0·17 (0·33)
	Non-operators	0·28 (0·68)	0·39 (0·65)	0·11 (0·61)

The same scores can be derived from the fourth trial in which the sewing machine was used. In this case, however, there are two other measures of performance available. These can be obtained from the tachograph records. The maximum machine speed during a response can be extracted. From these the mean maximum machine speed and the standard deviation of the maximum machine speeds during a trial can be calculated. All these scores and their standard deviations for this trial are shown in Table 2. It will be seen that, at this task, the machine operators, as compared with non-operators: (*a*) achieve a greater number of correct responses ($t=4{\cdot}15$, $p<0{\cdot}001$); (*b*) achieve a more consistent response ($t=3{\cdot}13$, $p<0{\cdot}01$); and (*c*) achieve a response which on average is more correct ($t=2{\cdot}93$, $p<0{\cdot}01$). The machine operators also achieve a higher and more consistent maximum machine speed, but these differences are not significant. Individual differences between members of the group are less in the case of operators. That is, all the operators have about the same score by each measure. Of the non-operators one or two are about the same as the operators, but other non-operators are poor by comparison.

Table 2

	4th Trial	
	Operators	Non-operators
Number of subjects	10	11
Number correct	20·4 (4·2)	12·5 (3·5)
Standard deviation	1·29 (0·10)	2·08 (0·71)
Mean	−0·09 (0·35)	0·76 (0·82)
Mean maximum speed obtained*	64·7 (7.9)	58·2 (10·7)
Standard deviation of maximum speeds	9·6 (1·7)	13·2 (4·0)

* The speed units are such that the maximum possible machine speed is 100 units.

The means and standard deviations of the various scores obtained during the three trials using the foot-controlled Morse key are shown in Table 3. There is no significant difference in any score between operators and non-operators. Comparison with Table 1 suggests that there is little difference between performance with the hand and with the foot.

Table 3. *Number in sample: ten operators; eleven non-operators*

		1st Trial (slow)	2nd Trial (fast)	3rd Trial (slow)
Number correct	Operators	14·0 (6·1)	12·8 (4·2)	17·6 (5·6)
	Non-operators	15·1 (6.8)	9·0 (7·1)	19·3 (6·4)
Standard deviation	Operators	2·07 (0·17)	2·00 (0·66)	1·79 (0·32)
	Non-operators	1·81 (0·58)	2·22 (0·69)	1·51 (0·50)
Mean	Operators	0·36 (0·84)	1·07 (1·32)	−0·33 (0·66)
	Non-operators	1·18 (1·01)	2·37 (1·33)	0·17 (0·52)

The final trial in this experiment involved hand control of the sewing machine. The results are shown in Table 4. In contrast to Table 2 it is quite clear that in this case there is no difference between operators and non-operators.

Table 4

	4th Trial	
	Operators	Non-operators
Number of subjects	10	11
Number correct	16·1 (3·7)	14·9 (3·5)
Standard deviation	1·62 (0·37)	1·85 (0·37)
Mean	−0·05 (0·53)	+0·25 (0·67)
Mean maximum speed obtained	74·7 (1·0)	76·9 (3·5)
Standard deviation of maximum speeds	6·8 (2·0)	8·0 (2·0)

V. CONCLUSION

No difference has been detected between machine operators and non-operators in ability to make a precise and consistent short response by: (*a*) pressing the finger on a key; (*b*) pressing the foot on a key; (*c*) using the hand to control a sewing machine. In ability to make a similar response using the feet in control of a sewing machine, machine operators are significantly better than non-operators.

VI. DISCUSSION

If this experiment is considered as a transfer situation, the pre-transfer task consists of about 20 years' experience in producing short runs of stitching within a more general skill. The experiment involves an experimental group (sewing machine operators) with previous training and a control group (non-operators) without previous training. Since both groups were exposed to the same sequence of experimental tasks, it seems feasible to neglect the transfer effects between these tasks. Thus there were effectively four post-transfer tasks. (i) The foot-treadle sewing machine. This involved using the same receptor and effector systems to make isolated responses which result in short runs of stitching (Table 2). (ii) The finger-controlled Morse key. This involved making isolated responses using different sensory cues and a different effector mode (Table 1). (iii) The foot-controlled Morse key. In this case the effector system is unchanged but the receptor system is different (Table 3). (iv) The hand-controlled sewing machine. The effector system is changed but the receptor system is only partially modified (Table 4).

The difference between the groups for (i) suggests that the element under investigation will transfer from a general serial skill to a more isolate situation. This is important since, as Bartlett (1948) has pointed out, the study of skill by the isolation of particular parts of the general task can be misleading. Confirmation that in this case the isolated response will transfer out of the serial skill is necessary before any further conclusions can be drawn.

In the introduction it was suggested that sewing machine-operators might have acquired a central mechanism appropriate to this specific response which would function independently of peripheral mechanisms. From (ii) there is no evidence to support the existence of such a mechanism. It remained possible, at this stage, that the mechanism as envisaged might be detected by leaving either the effector or receptor system unchanged. This was attempted in (iii) and (iv), but again there was no evidence of positive transfer. It seems then that this particular motor skill is very sensitive to changes of the receptor and effector modes.

This approach to the problem of transfer is related to the classical analytical method first developed by Poffenberger (1915) and Wylie (1919). These and later experimenters have concluded that useful generalizations concerning transfer can be made in terms of response and stimulus similarity. The change of terminology from stimulus/response to receptor/effector emphasizes the importance of considering how the subject acts rather than what the action achieves. The consideration of transfer in terms of stimulus and response has led to the relative neglect of the specific information channels which are utilized and the amount of 'coding' of information which is necessary to carry out the task effectively. Thus some experimenters have dealt with situations in which visual symbols are used to evoke a linguistic response (learning of nonsense syllables) while others have used visual displays which require merely a direct motor response. The same stimulus/response conceptual framework is applied. Yet in the former case the information is quite strictly channelled in and out with the highly practised visual and vocal systems so that any transfer effects which occur are almost certainly central in origin. In the case of motor skills where there is a great variety of sensory cues available it seems unlikely that the situation can be understood unless attention is paid to the particular ways in which the peripheral systems are being utilized.

To return to the experiments described above, it will be noted that many years have passed since the operators actually learned their pre-transfer task. It seems reasonable to assume that in this case we are dealing with the transfer of skill as such and not with the mixture of this and transfer of 'learning how to learn' which has confused the results in most experiments where all the transfer tasks are learnt in the laboratory within a short period. These experiments can be regarded as further evidence to support the hypothesis that the large transfer effects which are normally obtained in the laboratory are largely dependent on the transfer of learning technique rather than the utilization of particular stimulus/response mechanisms. This may be of some practical importance, since the present rate of technical change in industry is increasing the necessity for workers to change their skills. If the above hypothesis is valid they will not gain much assistance from positive transfer, but at least the old skill will not impede the learning of the new one.

The research was sponsored by the Department of Scientific and Industrial Research/Medical Research Council Joint Committee on Individual Efficiency in Industry, and financed from Counterpart Funds derived from United States Economic Aid.

REFERENCES

BARTLETT, F. C. (1948). The measurement of human skill. *Occup. Psychol.* **22**, 31–8.

POFFENBERGER, A. T. (1915). The influence of improvement in one simple mental process upon other related processes. *J. Educ. Psychol.* **6**, 459–74.

WYLIE, H. H. (1919). An experimental study of transfer of response in the white rat. As quoted in R. S. Woodworth & H. Schlosberg (1955), *Experimental Psychology*. London: Methuen.

(*Manuscript received* 11 *February* 1956)

A NOTE ON THE ROLE OF 'GUIDANCE' IN LEARNING

By J. M. von WRIGHT
Institute of Psychology, University of Helsinki

The traditional experiments on 'forced correct responses' as a method of guidance have indicated that its facilitative effect on learning is rather small. It is suggested that this outcome of the experiments was mainly due to the fact that the subjects during the guided trials got little or no information as to the location of the choice points or the possible choices at these. An experiment is described showing that when such information is provided, 'forced correct responses' may be a very efficient form of guidance. The importance of information about possible errors for the functional discrimination of the correct responses is briefly discussed.

I. Introduction

A number of studies have been concerned with the efficiency of various kinds of guidance in perceptual-motor learning. With guidance the learner does not discover the solution of a problem or the most efficient way of performing a task independently, errors being minimized by guiding the learner—wholly or partly—through the correct responses. In several experiments errors have been prevented altogether during a part of the learning period, the subject being forced to make only correct responses.

These experiments have mainly been carried out using stylus mazes. The typical procedure has been to guide the subject along the correct path from start to goal during some stage of learning and then to discontinue the guidance, leaving the subject to practice for himself until the criterion is reached. The main modes of 'forcing' the subject's responses have been (*a*) mechanical guidance, by blocking the entrances to the blind alleys (Koch, 1923); (*b*) manual guidance, the experimenter guiding manually the stylus along the correct path (Ludgate, 1923; Melcher, 1934; Wang, 1925; Waters, 1930); and (*c*) verbal guidance, the experimenter telling the subject what moves to make while traversing the maze (Wang, 1925). The results from the various experiments are in fairly good agreement, showing that guidance in most cases is favourable to learning, its relative effectiveness decreasing with an increasing amount of guidance, and decreasing the earlier it is inserted in the learning process (cf. Carr, 1930). The best results were obtained with a small amount of guidance during the very early trials. However, even under favourable conditions the facilitation due to guidance was found to be rather slight, and in no case could it to any major extent replace actually doing the task to be learned by 'trial and error'. For example, even when subjects were given forty or eighty initial guided trials on a maze which was learned on the average in 37·5 trials without any guidance, they still required (on the average) more than seventeen trials to reach the criterion when allowed to trace freely (Waters, 1930).

These results are still widely quoted without further comment in text-book chapters on the economy of learning (e.g. McGeoch & Irion, 1952; Wolfle, 1951). In view of this and of the importance of the issue for various training procedures, it may be of some interest to point out that an important limitation characterized the experimental procedure in all the studies cited above; during the guided trials the subjects were given little or no information as to the location of the choice points and the possible choices at these. The movements required to trace a maze are generally quite easy to perform once the maze is adequately structured perceptually. Usually the choice points in a maze serve as 'reference points' for the organization of the pattern to be learned. When clear-cut information about these points is omitted, it seems likely that 'forced' correct responses serve to facilitate the general orientation toward the goal as well as the building up of a general 'picture' of the true path, but a more detailed organization becomes difficult in the absence of such information.

Most experimenters in the field of maze learning agree that subjects preferably learn most types of mazes by a 'verbal' method, and that this generally is by far the most efficient way of dealing with a maze problem (cf. von Wright, 1955). Taken by itself, the correct path is mostly a rather complex affair and it is comparatively difficult to learn as a purely 'motor' sequence. But even those subjects who learn by a motor method do not normally learn to perform the sequence of movements along the correct path as a serial skill which may be correctly executed in the absence of appropriate external guiding stimuli. This is indicated, *inter alia*, by the fact that when subjects who have learned a complex maze are asked to draw the correct patterns on a blank sheet of paper on reaching the criterion, the reproductions made by subjects using a motor method of learning are on the average even more inaccurate than those drawn by subjects using a verbal method. Irrespective of the learning method employed, the absence of a fixed set of reference points will thus presumably make the task difficult to structure. The subject may, of course, try to learn all the characteristics of the path (many of which are irrelevant to the subsequent correct performance of the task), or he may act on the expectation that only some particular aspects of the path (e.g. the direction of the turns) are of importance, while disregarding other aspects such as the distances between the turns. However, if information as to the choice points and the possible choices is supplied, the task is much simplified in that the subject knows what characteristics of the path are the relevant ones, and can concentrate on learning only these. In other words, the range of possibly relevant information is considerably reduced. In addition to this, such information presumably increases the discriminability of the correct responses, as these are 'reinforced' as being *the correct ones out of sets of specified alternatives*.

It is thus suggested that 'forced' correct performance of a maze task is a fairly inefficient form of guidance as long as the task cannot be adequately structured perceptually (or verbally), that the conditions in the traditional studies of guidance were rather unfavourable to this 'structuring', and that giving the subjects information about the choice points during the guided trials will considerably increase the efficacy of guidance. Some experimental data obtained by the writer are relevant to these suggestions.

II. Experimental apparatus and procedure

A detailed description of the apparatus and general procedure used has been reported elsewhere (von Wright, 1955). Briefly, patterns consisting of twelve diamond-shaped units separated by short straightways were drawn on strips of paper and repeatedly presented to the subjects on a revolving drum. The subjects traced the moving pattern with a stylus through an aperture in a screen placed in front of the drum. The aperture was 2 cm. wide, permitting only a part of a unit to be seen at a time. All the choices were made from the same spatial point, and (as the speed of presentation was constant) at equal intervals. The time interval between any two successive choices was 1·75 sec., the total time per trial being 21·8 sec. Between successive trials there was an interval of 10 sec. The split-half reliability of the task, errors on odd *v.* even trials (guided trials omitted), exceeded in each case 0·92 (uncorrected). The correct pattern was *LRRLRLLLRRLR*, and the learning criterion was two successive errorless trials.

Sixty undergraduates, unfamiliar with learning experiments and with the problem studied, served as subjects. They were randomly divided into three groups. In group A (control) the subjects learned throughout by the anticipation method. When a choice had been made, information about its correctness was given by two red lines drawn across the wrong path. Both experimental groups received *four initial guided trials*, after which they immediately continued to learn the task by the anticipation method under conditions identical to those of group A. In group B (guidance 'with information') the subjects were given full information during the guided trials about the location of the choice points and the possible choices at these. This was achieved by using the same pattern that was presented to the control group, but drawing the red lines indicating wrong choice close to the choice points where they could be seen prior to the moment of (overt) choice. In group C (guidance 'without information') the subjects were presented, during the guided trials, with the correct path only. The differences between the three conditions can be seen from Fig. 1, in which is shown a portion of the pattern as presented to each of the groups during the four first trials. After the fourth trial, the pattern 'A' was presented in a similar manner to all subjects. Non-correction procedure was employed throughout. The instructions were the same for all groups, apart from an additional clause in those given to groups B and C, in which the subjects were told that they (in group B) would be forced to make only correct choices during the initial trials, or (in group C) would be presented with the correct path only during these trials. No information about the shape of the pattern was given in the instructions.

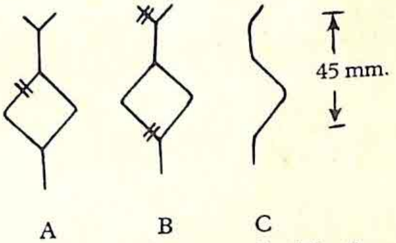


Fig. 1. A part of the pattern as presented under each of the three experimental conditions.

Results

The mean trial scores and error scores for each group are presented in Table 1. For groups B and C the trial scores are given separately for all trials and for the unguided ones only. From the error score of the control group (A) the errors made during the four first trials have been excluded in order to make the three groups comparable, as no errors could be made during the guided trials in groups B and C. (The mean total error score for group A was 98·6.) The time scores are not presented, as they—the task being paced—are strictly parallel to the trial scores.

In all scores group B (guidance 'with information') was very much superior to the other groups, while C (guidance 'without information') was to a less extent—but still significantly—better than the control group. Forced correct responses 'with information' could in fact wholly replace any trial and error learning; out of ten subjects who were given twelve guided trials under condition B, eight were subsequently able to perform the task by the anticipation method (condition A) without making a single error. The results are thus in agreement with the prediction.

Table 1. *Mean trial scores and error scores*

	Trials			Errors (after fourth trial)	
	Total		Unguided		
Groups	*M*	σ	*M*	*M*	σ
A	24·25	7·61	—	76·30	37·17
B	10·45	5·22	6·45	15·80	15·38
C	18·80	4·13	14·80	52·15	26·92
Differences					
A − B *t* =	6·52			6·56	
p =	0·0001			0·0001	
A − C *t* =	2·74			2·29	
p =	0·01			0·05	
B − C *t* =	5·47			5·11	
p =	0·0001			0·0001	

III. DISCUSSION

Under the conditions of the present experiment forced correct responses facilitated learning much more when information about the choice points and the possible choices at these was provided than when it was omitted. In the latter case the subjects only got 'undifferentiated' information about the correct responses, and had difficulties in identifying these responses later when they appeared in the context of a set of fairly similar alternatives. Presenting the correct alternatives as contrasted with the wrong ones served to increase their subsequent identifiability, and provided the subjects with clear-cut points of reference.

Although the experiments discussed have been concerned only with maze learning, the principles involved seem to have a more general applicability. A number of tasks are similar to the maze task in consisting of part-activities which in themselves are fairly easy, the main difficulty being to discriminate functionally between correct and incorrect alternatives at successive points of the performance. Here information about alternative 'wrong' cues (or possible wrong responses) while the subject is guided to perform the task correctly serves to indicate which are the dimensions of variation to which the subject should learn to respond differentially (or where the response should be a particular one within a set of similar responses and not any one of this set). Looking at the problem more broadly, such information may, in a variety of situations, serve to facilitate the discrimination of 'cue' stimuli. Thus it was found by the writer in experiments on concept formation in children and the complex visual discrimination learning in adults, that the more extensive the information about possible 'relevant' errors, the more easily were the correct cues discriminated from any *new* alternative wrong ones. However, the available data do not permit any extensive generalizations, and there are some obvious limitations to the usefulness of 'forced correct responses with information', in the present sense, as a method of teaching, in particular when applied to the teaching of motor skills (cf. Welford, 1951).

REFERENCES

Carr, H. A. (1930). Teaching and learning. *J. Gen. Psychol.* **37**, 189–218.

Koch, H. L. (1923). The influence of mechanical guidance upon maze learning. *Psychol. Monogr.* no. 147.

Ludgate, K. E. (1923). The effect of manual guidance upon maze learning. *Psychol. Monogr.* no. 148.

McGeoch, J. A. & Irion, A. I. (1952). *The Psychology of Human Learning.* Rev. ed. New York: Longmans, Green.

Melcher, A. W. (1934). Children's motor learning with and without vision. *Child Developm.* **5**, 315–50.

Wang, T. L. (1925). The influence of tuition in the acquisition of skill. *Psychol. Monogr.* no. 154.

Waters, R. H. (1930). The influence of large amounts of manual guidance upon human maze learning. *J. Gen. Psychol.* **4**, 213–27.

Welford, A. T. (1951). *Skill and Age.* London: Oxford University Press.

Wolfle, D. (1951). Training. In Stevens, S. S. (ed.): *Handbook of Experimental Psychology.* New York: Wiley.

von Wright, J. M. (1955). An experimental study of human serial learning. D.Phil. Thesis, University of Oxford.

(*Manuscript received* 16 *April* 1956)

ATTENTION TO CHANGE, CONDITIONED INHIBITION $({}_{S}I_{R})$ AND STIMULUS SATIATION

By D. E. BERLYNE

Psychology Department, University of Aberdeen

Berlyne (1951) reported an experiment on attention to change. Human S's were first subjected to a habituation phase, in which they had to make key-pressing responses to a series of homogeneous visual stimuli. They then underwent the experimental phase, in which they had to make choices between responding to stimuli of the same sort as before and responding to others of a different shape or colour; they showed a tendency to respond more frequently to the *novel* than to the *familiar* stimuli.

A further experiment has since been performed with a view to investigating some of the questions left unanswered by the previous study. In the experimental phase of the latter, the novel stimuli always appeared at the same place for a given S. It is possible, therefore, that he kept watch on that place in a way which accounts for the greater probability of responding to novel stimuli. One purpose of the present experiment was to show whether attention to change occurs when novel and familiar stimuli may appear at the same places, in which case a preference for responding to novel stimuli must be due to processes (which may include receptor adjustments) coming into play after the stimuli have been presented.

Another object of the experiment was to test predictions from two possible explanations of the phenomenon. Berlyne (1951) hypothesized that Hull's 'conditioned inhibition $({}_{S}I_{R})$' was responsible. This is a learned and hence permanent inhibition affecting a particular S–R bond. On this hypothesis, the attention-to-change effect should be just as strong if 24 hr. elapse between habituation and experimental phases. An alternative concept that might be invoked is 'stimulus satiation', introduced by Glanzer (1953) as a factor assumed to underlie the rat's spontaneous alternation in the T-maze and taking the form of a reduced tendency to make any response to a recently exposed stimulus. A similar temporary stimulus-specific inhibition (to use Hinde's (1954) terminology) figures in Broadbent's theory (1953). Attention to change, if it is due to such a factor, should be just as strong if the responses performed in the habituation phase differ from those performed in the experimental phase.

PROCEDURE

The S sat facing a vertical wooden board, about 4 ft. 6 in. away, with four 2½ in. square apertures, numbered 1–4, in a horizontal line at about eye level. A circular spot of red or white light, *c.* ¾ in. in diameter, could appear in the middle of each aperture. A Palmer timing unit made such stimuli appear for 5 sec. with 5 sec. intervals. S's right hand rested on a board with four Morse keys, numbered 1–4 and arranged in an arc.

There were thirty S's, taken from the first- and second-year psychology classes, and divided into three groups of ten as follows.

(1) *The immediate–same* (*IS*) *group* first underwent a habituation phase with stimuli of one colour—red for five S's (the 'red subgroup') and white for the remaining five S's (the 'white subgroup'). A random sequence of sixty stimulus patterns was presented. Of these, the first and second thirty consisted of twelve single stimuli (three in each of the four positions) and eighteen stimulus pairs (three in each of the six possible pairs of positions). When a single stimulus appeared, S was to press its corresponding key. When two stimuli appeared, he was to respond to either but not both, in fact to the one he noticed first.

Immediately after the habituation phase followed the experimental phase, in which both red and white stimuli were used. Eighty stimulus patterns were presented, of which each successive twenty was made up of eight single stimuli (four of each colour distributed equally among the four positions) and twelve pairs consisting of a red and a white stimulus (each of the twelve possible combinations of positions was used once). The task was as before.

(2) *The immediate–different* (*ID*) *group* had the same procedure, except that a different task was imposed for the habituation phase. S was told that either one or two lights would appear at any one time, and he was instructed to voice a guess during each interval with respect to the number that would appear next. In the experimental phase, which followed immediately after, the key-pressing task was used.

(3) *The 24 hour–same* (*24S*) *group* had exactly the same treatment as the IS group, except that approximately 24 hr. elapsed between the two phases.

Results

In the IS group's experimental phase, the mean number of choices in favour of red stimuli was 20·4 for the 'red subgroup' and 26·4 for the 'white subgroup', as compared with 24 to be expected by chance. According to a one-tailed Wilcoxon test, $T = 16$ and $p < 0{\cdot}01$. The colour used in the habituation phase seems, therefore, to affect the choices made in the experimental phase. A tendency to respond to novel rather than to familiar stimuli is thus confirmed in the conditions of the present experiment and may consequently be attributed to factors occurring after the onset of the stimuli.

The mean number of choices (out of forty-eight) in favour of novel stimuli made by the IS group (27·0) is significantly greater than the mean numbers made by the ID group (21·0) and the 24S group (22·9): $t = 2{\cdot}10$ and $2{\cdot}24$ respectively, d.f. = 18, and $p < 0{\cdot}05$ in both cases. The hypotheses that a 24 hr. interval or a change of response between phases will not diminish the attention-to-change effect can thus be rejected, and explanations in terms of ${}_{S}I_{R}$ or stimulus satiation *alone* are not borne out.

The failure to find support for the stimulus-satiation theory is interesting in view of several experiments on maze alternation in the rat (Walker, Dember, Earl, Fliege & Karoly, 1955; Kivy, Earl & Walker, 1956; Dember, 1956) which cast doubt on its generality. The rat appears to turn away from stimuli which it has recently encountered *in the same context*.

REFERENCES

BERLYNE, D. E. (1951). Attention to change. *Brit. J. Psychol.* **42**, 269–78.

BROADBENT, D. E. (1953). Classical conditioning and human watchkeeping. *Psychol. Rev.* **60**, 331–9.

DEMBER, W. N. (1956). Response by the rat to environmental change. *J. Comp. Physiol. Psychol.* **49**, 93–5.

GLANZER, M. (1953). Stimulus satiation: an explanation of spontaneous alternation and related phenomena. *Psychol. Rev.* **60**, 257–68.

HINDE, R. A. (1954). Changes in responsiveness to a constant stimulus. *Brit. J. Anim. Behav.* **2**, 41–55.

KIVY, P. N., EARL, R. W. & WALKER, E. L. (1956). Stimulus context and satiation. *J. Comp. Physiol. Psychol.* **49**, 90–2.

WALKER, E. L., DEMBER, W. N., EARL, R. W., FLIEGE, S. E. & KAROLY, A. J. (1955). Choice alternation: II. Exposure to stimulus and place without choice. *J. Comp. Physiol. Psychol.* **48**, 24–8.

(*Manuscript received* 17 *April* 1956)

A NOTE ON THE FALLACY OF THE 'PHENOMENOLOGICAL FALLACY'

By J. R. SMYTHIES

Psychological Laboratory, University of Cambridge

In a paper in this *Journal* entitled 'Is consciousness a brain process?' Place (1956) claims that physiologists such as Sherrington, who have recognized the existence of mental (or psychical) events *and* physical events, have committed 'a fairly simple logical error' which Place calls 'the phenomenological fallacy'. This mistake consists in supposing that 'when the subject describes his experience, when he describes how things look, sound, smell, taste or feel to him, he is describing the literal properties of objects and events on a peculiar sort of internal cinema or television screen, usually referred to in the modern psychological literature as the "phenomenal field"'. The fallacy concerned is 'the mistaken assumption that because our ability to describe things in our environment depends on our consciousness of them, our description of things are primarily descriptions of our conscious experience and only secondarily, indirectly, and inferentially descriptions of the objects and events in our environments'. Place says (1) that we do not describe things by first describing their phenomenal properties and then inferring their real properties from their phenomenal properties, and (2) that we learn to recognize the real properties of objects before we can describe our conscious experience of them. We describe our conscious experience not in terms of phenomenal properties but, 'by reference to the actual physical properties of concrete physical objects, events and processes which normally, though not perhaps in the present instance, give rise to the sort of conscious experience which we are trying to describe. In other words when we describe the after-image as green, we are not saying that there is something, the after-image, which is green, we are saying that we are having the sort of experience which we normally have when, and which we have learned to describe as, looking at a green patch of light' (p. 49).

Now there are a number of entangled confusions and mistakes in this argument which I shall attempt to unravel. Everyone would agree that it is false to say that when a man is describing some physical object it is really his own experience, or his own sensations, that he is describing. But we cannot conclude from this that a man can never describe his own sensations, or that when he is describing his own sensations he is really describing physical objects. For it is true to say that those professionally concerned with the analysis of sensation can take up what Quinton (1955) has called the phenomenological frame of mind and can use words to denote their sensations and can thus describe their sensations with no necessary reference whatever to physical objects. Furthermore, even when a man is describing a physical object he is certainly also having sensations, which he could have described instead of the physical object, if he had adopted the special phenomenological frame of mind and used the 'sensible object' language (or the 'sense-datum' language) instead of the physical object language. Now it is perfectly true that the physical object and the sensible object languages contain many words in common (e.g. 'red', 'round', 'ticklish', etc.), but these words are used in each language in perfectly specifiable ways. For instance, 'this sense-datum is red' describes a fact of direct sensory experience; 'this material thing is red' may be used in a situation where a certain physical object evokes red sense data in the visual field of people under certain specified physical conditions (e.g. the light must be white or red and not blue or green, the person must not have taken santonin, etc.). It is also true that these words have been learned in the context proper to the use of the physical object language and not the sensible object language, for we usually teach children the use of words by showing them physical objects. But this fact has no logical significance, for it is only contingent as the following examples show.

(*a*) A mother could always teach her children the names of colours by reference to after-images and not objects. She could ask her children to look at coloured lights and could say of each after-image as it appeared 'that's *blue*', 'that's *green*', etc. The child could then use these words to refer to the colours of physical objects. Furthermore, a child could learn the proper use of 'green', for example, without ever having seen any green light or green physical object. All we would have to do would be to bring the child up in a red light, and then at a suitable age ask the child to look at a bright red light and say of the resultant after-images in the complementary colour 'that's *green*'. It seems plausible to suggest that after this instruction the child should be able to name correctly any green physical object shown to it in a white light. In which case Place would have to say that whenever the child subsequently looks at a green object and describes it as green, he is *only* having the sort of experience which he had learned to describe as green when looking at a green after-image. So his argument against the thinghood of after-images cannot be upheld on these grounds.

(*b*) Secondly, can we doubt that had hallucinations been very much commoner than they are, and many of them collective (as some are even now), that in this case the use of many words would have been obtained in their sensible object context and not their physical object context?

Now it is clear that the question as to whether an after-image is a thing (or an entity) or not depends on what we mean by 'thing' (or an entity), and we can construct our definition of 'thing' so as to include or exclude after-images as we wish. However, it is important to ask if a scientific explanation of perception should contain 'after-image' as a basic term. In other words do we *observe* after-images? *Do* after-images have the properties they appear to have, i.e. spatial extension, causal relations and colour? Why, if an after-image is not a thing, is it so confoundedly thing-like? And if the answer to the first two of these questions is 'No', what is going on when a person claims to be observing an after-image? A behaviourist account would run as follows: 'When a man says he is observing an after-image, all that is happening is that a certain excitation is reverberating in his retinal and visual cortex leading to correlated excitation of the motor speech centres via the computing and communicating mechanism connecting the two. Thus certain noises are made that may be depicted by the symbols "I am seeing an after-image". On the other hand, no such utterance may be made at the time, but a memory trace of the reverberating activity may be laid down so that a subsequent statement "Last week I had an after-image" may be made.'

The behaviourist would claim that the after-image does not have the properties it appears to have, for all we are dealing with are the verbal responses of organisms which use a misleading 'object' language when their outputs are activated by certain internal events in their mechanism. The organism should not issue the symbols 'I am observing an after-image', or even (vide Place) 'I am now having the sort of experience that I would have had had I looked at such and such a light', but only 'I am now making the sort of statements that I would have made had I been observing such-and-such a light'. The alternative account states that images are a part of the fundamental furniture of the mind and do have the properties that they appear to have.

Now it is not necessary to discuss here which of these two accounts is *correct*. I only wish to point out that neither makes any logical mistake. They are equally valid ontologico-empirical statements. One says there are only physical objects. The other says there are physical objects *and* images. The only way in which we could possibly decide between them would be by experiment. I have suggested elsewhere (1956) how this problem could be approached.

Now Place says that an after-image is not merely what a red light of a certain kind would have looked like had we seen it, but that in the case of the after-image we are having an experience similar to that which we would have had if we had looked at a certain light. To this we can say that it is difficult to see to what this term 'similar experience' refers in such a usage. 'Experience' usually refers to *events* experienced. For example, in the common-sense usage, 'I had a similar experience ten years ago' means that ten years ago I observed and lived through certain events of similar kind to those under consideration; or, in the technical phenomenological usage, it means that certain phenomenological (experiential) events occurred in my sense fields ten years ago.

But it seems odd to say that we could have an experience of *nothing*. Surely to experience nothing is not to have an experience? Surely a sensory experience is nothing more than a certain group of events, i.e. the causal processes of perception both outside and inside the organism in the more extended use of 'experience' ('I have had some interesting experiences in my time') and certain experiential events in the technical phenomenological use of 'experience'. It would surely be a mistake to say that we are dealing with the causal chain *and* the experience, so that we could logically have no events and *yet* have an experience. For this is clearly the same category mistake that Ryle claims the Cartesian dualists make over 'mind'. To this Place might reply that, in the case of the after-image, certain cerebral events were going on to cause the experience. But surely it is only clumsy and misleading to speak of cerebral events, or of events in the perceptual mechanisms, as causing (or being correlated with) the 'experience of nothing' when the experience is nothing more than the physiological causal chain. The physiological causal chain cannot cause the experience—it *is* the experience. In which case, we do not have to use the language of experience at all but only the language of the radical behaviourist theory.

Thus we can say that it is certainly false to say that, when people are describing physical things, they are really describing their sensations. But it is also false to say that, when people are describing their sensations, they *must* really be describing the physical things causally related to their sensations, or the sort of things that would be causally related to their sensations, and so on. Furthermore, it is a contingent fact that the use of words is obtained by reference to physical objects and we can easily imagine at least some instances where this would not be so.

So we may conclude that the statement 'there are after-images and they are green entities' is not derived from any logical fallacy but it merely expresses one particular ontological theory about images, and the statement 'there are no after-images' merely expresses another ontological theory about images and neither can be proved or disproved on logical grounds alone.

I am most grateful to Prof. O. L. Zangwill, Mr James Thomson and Mr Richard Gregory. To them are due whatever backbone or finesse this note may have.

REFERENCES

PLACE, U. T. (1956). Is consciousness a brain-process? *Brit. J. Psychol.* **47**, 44–50.
QUINTON, A. M. (1955). The problem of perception. *Mind*, **64**, 28–51.
SMYTHIES, J. R. (1956). *Analysis of Perception.* London: Routledge and Kegan Paul.

(*Manuscript received* 21 *April* 1956)

PUBLICATIONS RECENTLY RECEIVED

A Study of Thinking. By J. S. BRUNER, J. J. GOODNOW and G. A. AUSTIN. New York: Wiley; London: Chapman and Hall. 1956. Pp. xi+330. 44*s*.

This book must be reckoned one of the most important contributions yet made to the study of concept formation. It breaks away completely from the problems which dominated investigation into this subject in the twenties and thirties (e.g. whether concepts can be formed unconsciously, what sort of images accompany the formation of a concept, etc.); such problems have revealed their sterility, and have been given a decent burial in the pages of Humphrey's book on 'Thinking'. The present authors are of course not the only workers in the field who are employing a new approach to the study of concept formation; the book is very much the product of the *zeitgeist*, and draws heavily on concepts developed in games theory and information theory.

The question to which this book attempts an answer is 'How do we form concepts?' The authors have gone some way towards giving a new answer to this problem by taking two important steps. First, they use the rudiments of games theory to analyse what methods we might expect to be used in forming concepts and how these methods might be expected to vary with variations in the conditions under which concepts might be formed. The methods used to arrive at concepts are termed 'strategies'. Secondly, in the experiments reported, the authors have developed greatly simplified experimental techniques which enable them to analyse the strategies with which their subjects are operating.

The easiest way of illustrating these two points is to outline part of the experimental technique. An array of sixty-four figures was set out before the subjects; it was explained to the subjects that each figure exemplified one of two possible values of six attributes, and subjects were shown what these attributes and their values were. Subjects were then told that they had to form a concept; figures which were an instance of that concept could be defined by certain common attribute values. All subjects were then shown a positive instance of the concept, and told to arrive at the concept by choosing further figures to see whether they were or were not instances of the concept. It should be noted that the great advantage of the experimental technique used is that, because there are only a finite number of possible concepts and all these are made clear to the subject before the start of the experiment, it is possible to perform the following types of analysis on the data: one can say how much information a subject can acquire by any given choice and compare this with the amount of information he makes effective in his guesses; knowing the choices a subject has made, it is possible to say what further choices would extract the maximum amount of information from untested figures and to see how far a subject's actual further choices satisfy the condition of obtaining the maximum amount of new information; it is also possible to state how a subject's choices depart from random choices.

The authors give a subtle analysis of factors likely to influence the pattern of successive choices; thus, in addition to regulating the risk of failure to attain the correct solution in a given number of choices, they point out that certain strategies can facilitate the task of assimilating and storing information (e.g. considering only one hypothesis at a time and relating information received only to that hypothesis), and they are able to show that a strategy which makes the assimilation of information easy is proportionately more successful as the experimental conditions render it more difficult to assimilate and retain information (e.g. they compare conditions where subjects can see the results of all past choices with conditions where they can no longer see the results of past choices but must store the information in their head if they wish to use it).

As well as investigating how subjects select figures to attain a conjunctive concept, the authors have investigated the strategies used when subjects are given no choice of the order in which positive or negative examples of the concept are presented to them; strategies are revealed by forcing the subjects to write down what they think the concept is after each example has been presented. They have also investigated the attainment of disjunctive concepts, i. e. the formation of a concept consisting of a disjunctive combination of attribute values. A further major novelty in their approach is that they attempt to investigate categorizing where the cues to be used in the categorizing are only probabilistic. This, of course, has been done before, but the authors try to give a rationale for matching event probability in frequency of guessing the event, and to show under what conditions such matching takes place.

A large number of variables affecting the kind of strategy which will be adopted, and the difficulty of correct concept formation are subjected to theoretical analysis in terms of the concepts of games theory, and many ingenious experiments are reported to test predictions arrived at in this way. It is clear that the authors and other recent workers in this field have opened up an entirely new range of problems, and a very promising field for further research. The authors make it clear throughout that the type of theoretical analysis they are employing opens up far more problems than they have attempted to solve; the important point is, that by developing a new way of analysing the problem of concept formation and new experimental techniques, they at the same time raise new problems and provide a method for solving them. In a way, the sort of explanation given throughout the book is like the explanations of behaviour given in everyday life—the main explanatory concepts employed are aims and expected outcomes. The virtue of games theory is that it has provided a technique for the quantification of such concepts, and offers the psychologist some hope of being able to give this type of explanation and at the same time make more or less rigorous predictions. This book is the first large-scale attempt to apply this type of analysis to individual human behaviour, and to test the extent to which the facts of human behaviour can be fitted into the new conceptual framework; as such it is a portent of things to come. It suggests a way of dealing with those areas of behaviour which are too complex or about which too little is known to make it worth while casting hypotheses about them in a mechanistic or physiological form.

The book has only one serious fault: the exposition is not always as clear as it might be, partly because the writing is often clumsy and partly because attention is distracted away from the main argument by a series of comparisons with concept formation in everyday life which are often more misleading than helpful. For example, the difference between those who believe in localization of function in the brain and those who do not is far too complex for a comparison between the thought process of the two schools to form a good and helpful illustration of what the authors mean by wholist and part strategies.

There is a 65-page appendix to the book written by R. W. Brown in which he attempts to show that the way of looking at concept formation developed in the body of the book can be applied to the formation of linguistic concepts and to the formation of concepts in the world to which words apply. Though he gives an admirably clear account of some of the psychological problems to do with language, the connexion between his account and the main body of the book remains tenuous, and little effort is made to show what light is thrown on the problems of language by using the theoretical analysis and experimental techniques developed in the remainder of the book.

N. S. SUTHERLAND

Perception and the Representative Design of Psychological Experiments. By EGON BRUNSWIK. Pp. xii+154. University of California Press; London: Cambridge University Press. 1956. 37*s*. 6*d*.

It is some slight consolation for the loss of Egon Brunswik that he should have left behind this comprehensive statement of his views. No one interested in psychological theory can afford to ignore it, even though at times it may make unusually heavy demands upon his penetration and patience. For Brunswik could only express his ideas in the abstract idiom of a German philosopher, and the ideas themselves are not easy. His last sentence may serve to illustrate his style: 'Preception, then, emerges as that relatively primitive, partly autonomous, institutionalized, ratiomorphic subsystem of cognition which achieves prompt and richly detailed orientation habitually concerning the vitally relevant, most distal aspects of the environment on the basis of mutually vicarious, relatively restricted and stereotyped, insufficient evidence in uncertainty-geared interaction and compromise, seemingly following the highest probability for smallness of error at the expense of the highest frequency of precision.' But abstract and clumsy though this sentence may seem, it is not the work of a man whose primary concern is to expound a theory. Rather it is an effort to present the many-sided nature of the perceptual process as it reveals itself to a genuinely empirical study. It is we who become theoretical when we simplify our data in Lockeian or Gestalt terms.

The book is divided into two parts, of which the first was printed in the *Berkeley Symposium on Mathematical Statistics and Probability* in 1949. The second part, entitled *Perception, the Ecological Generality of its Distal Aim*, does little more than develop and illustrate some of the earlier ideas.

The argument of the whole book is based upon the account of perception which has been quoted above, If the sentence is studied adjective by adjective, it will be seen that a very wide range of experimental material is being taken into account. Ames, Bruner, Erismann, Gibson, Helson, Lashley, Tolman and Werner are among those whose work is relevant.

Now, if perception is the sort of thing it seems to be, then how can we study it? Brunswik provides a new analysis of experimental method in psychology, and shows the defects of, on the one hand, the traditional experiment with its manipulation of a single variable, and, on the other, the multifactorial approach, using analysis of variance in situations which have been artificially randomized. If perception were a purely causal affair (if, that is to say, stimuli *made* us perceive instead of *enabling* us to perceive), then the first kind of approach would be appropriate, as, indeed, it still is in sensory physiology. If the chances and events of this world came along as systematically as the plots of an agricultural experimenter, then we have statistical techniques which we could use. But perception and behaviour, and indeed the whole process of living, seems to fall somewhere between these two extremes of rigidity and randomness, and, on Brunswik's view, we need a third method to handle the sort of data prevalent in biology. In spite of some ingenious approximations, Brunswik admits that, although we can point to the sort of method we need, we do not yet have the mathematical tools which would form an essential part of it.

The further stages of the argument may seem open to criticism. If, as would seem to be the case, our perceptual world is an achievement made possible by cues which can deputise for one another and which may be differently weighted in every concrete preceptual situation, then the task of designing a 'representative' experiment appears to be impossible. Indeed, the whole distinction between an experiment in which the variables are known and controlled, and a real-life situation, tends to become blurred. Even allowing that we must sample a far wider range of possibilities than those normally present in a psychological laboratory, the decision as to whether a particular experiment can be regarded as 'representative' seems to remain, to some extent, arbitrary. What we have instead is an observational approach, using measurement where possible, but producing in the end a functional and descriptive account of perception rather than a causal analysis. It may be that this is what we should aim at, and it is certainly something worth doing from the point of view of biological science as a whole. But one is reluctant to accept the claim that this is the only way in which perception can be studied.

One may gain much from this book without accepting its general thesis. The acuteness of Brunswik's analysis and the obstinacy of his struggle with fundamental issues arouses admiration. If the text may sometimes seem perversely complicated, one is never quite sure that it is the writer who is to blame rather than his data. It is difficult to see how any future psychology of perception can avoid dealing somehow with the issues which Brunswik has raised.

JAMES DREVER

Social Structure and Personality in a City. By O. A. OESER and S. B. HAMMOND (Eds.). London: Routledge. 1954. Pp. xxii+344. 30*s*.

This is the first of two volumes edited by Oeser and Hammond. The second volume, to be reviewed later, describes social structure and personality in a rural community. The studies were undertaken by the Department of Psychology at Melbourne University from 1947 to 1950 as part of an international study of communities and social tensions commissioned by UNESCO.

In the preface of this volume, Prof. Oeser explains how successive class projects carried out by the Collective Behaviour course were used as a basis for much of the material and gave leads for conceptualization. These studies were considerably enlarged and systematized when the UNESCO grant became available. In the first chapter, the authors explain the orientation of the book: '...we have chosen to deal with topics concerned with the behaviour and outlook of individuals in a life context'. Historical, economic, ecological or political facts have been mentioned only in so far as they affect the individual. Primarily 'the inner, partly autonomous, psychological structure and dynamics of individuals' has been studied, 'only in so far as it affects their common every day living and beliefs. We are concerned with the determinants of social behaviour rather than with its remote or intrapersonal conditions.' The problem with which the writers deal has 'three aspects', i. e., 'the enunciation of sociological facts, the enunciation of psychological facts, and the investigation of the way in which the psychological facts rest upon the sociological'.

The book is divided into four sections. The first describes the Australian's attitude to his own and to other countries, his attitude to immigration and, in particular, his attitude to Jews. The description is based upon data obtained from Gallup polls and from studies carried out by the department. One of the studies deserves special mention, namely, that of relating the perception by Jews of the existing degree of anti-Semitic feeling to that degree of feeling actually elicited from non-Jewish respondents. This section is carefully planned and the discussion based on a wide range of data. It suggests useful leads for future inquiry.

The second section deals with the interaction pattern of the members of the family. It is based on a highly elaborate conceptual framework, confusing to the reviewer at least, and rendered more so by the topological models used. To test this conceptual framework, 128 11–12-year-old school children are asked in a group test to note which member of the family usually does each of thirty-three activities listed: 'Who usually does each thing, who helps, and who would never do it.' They then go over these activities again and are asked the following: 'It is quite likely that your parents do not always agree about all these things, so we want you to tell us how often they disagree.' The child is given a choice of categories—Often, Sometimes, Never—which he is asked to tick. The items range from 'Who decides when broken things should be repaired' and 'Which of your parents decides when to visit your mother's relatives' to 'Who decides how much money should be spent during the holidays'? No explanation is given for the selection of these activities, no doubts raised as to whether, for example, the child is really in a position to know who decides how much money is to be spent during the holidays.

On the basis of the answers to this group test and with no further check by relating these answers to those obtained from parents concerning these activities, the families are divided into four groups: husband-dominant, wife-dominant, autonomic (where each marital partner has autonomous areas of control) and finally the syncratic co-operative household in which husband and wife act and decide together. Further, each family is given a tension index based on the number of times the child has ticked that his parents sometimes or often disagree. These data in turn are related to age at marriage, occupational status, as well as signs of early maladjustment in the child (obtained from interviews with the mother) and to present signs of maladjustment in the child. A tetrachoric correlation of 0·078 between maladjustment and type of family structure is reported. No doubts are expressed concerning the validity of this unusually high positive relationship.

Some of the findings are both interesting and important, but to assess their value, further research would have to be done with techniques which depend less upon the child's perceptions of the situation. It is a pity that no mention was made in this work to the relevant studies carried out by Fels or Symonds.

The third section deals with the child and is based on tests given to the 128 children and on interviews with the parents of these children as well as on information collected from Kindergarten teachers on early training. These studies examine the behaviour expected of the child, and the techniques used in training, as well as the different roles which boys and girls are called upon to play.

The school children are questioned on the same areas of behaviour, using Bavelas Home Ideology questionnaire. Very interesting findings emerge, such as the relative degree of emphasis on conformity expected from boy and girl, the space of free movement allowed and the types of controls exerted. It would have been interesting, had time allowed, to relate the information obtained from parents to that obtained from the child. On the basis of TAT stories, the children were rated in terms of their adjustment to the family, a score which was then correlated to peer adjustment, to family tension and to family structure.

The fourth section deals with the place of work in social life, with class and family, with the relation of class membership to child training and with the effect of the individual's perception of his class membership on his behaviour and on his views on a variety of issues. This section is by far the best in the book. It shows the value of the multidimensional approach that has been adopted, namely, to collect a larger amount of data about the same families which makes it possible to determine the interdependence of attitude and behaviour. References to such obviously relevant studies as those of Havighurst and Davis are lacking.

The book is undoubtedly valuable. It discusses novel techniques of inquiry and attempts conceptualization. It is also perhaps rash and somewhat uncritical of the instruments used. Nevertheless, despite these deficiencies, perhaps inevitable in a pioneer work, it will occupy a significant place among the studies of child training and family life.

HILDE HIMMELWEIT

The Social Problem of Mental Deficiency. By N. O'Connor and J. Tizard. London: Pergamon Press. 1956. Pp. viii + 182. 30*s*.

The main purpose of this book is to give a summary of the research carried out over the last eight years by members of the Medical Research Council's Social Psychiatry Research Unit. Drs O'Connor and Tizard have, however, done much more than this in presenting the results of surveys and investigations within a framework of carefully collated statistical material culled from public and departmental reports. The outcome is a very clear and well-written account of the present status of certified mental defectives in England, and there is no other book where this factual material is so readily available or so clearly summarized. The title, however, is not an entirely happy choice, for, as the authors admit, their studies have been concerned more with occupational than social problems, and further information about the position of the mentally deficient person in the community is greatly needed.

The researches have usually involved two main stages: first, a broad survey of the particular field, and secondly, detailed experimentation upon a salient aspect of it. This has proved most fruitful. Unlike the medical investigations of the pathological aspects of mental deficiency, social surveys have, with the exception of the classical pre-war work of Burt, Lewis and Penrose, hardly been touched upon. In view of the radical changes in social and economic conditions following the war, a reassessment of this field was greatly needed. The book gives an account of the prevalence of mental deficiency and the services available to deal with the problem. Psychometric surveys yielded the surprising information that about half the largest clinical group, the feeble-minded, had I.Q.'s over 70 points on standard intelligence tests; this indicates that the legal concept of mental deficiency has widened since the time of the Wood Report. Somewhat unexpected results also emerged from experiments on learning in imbeciles, and these have important implications for occupation and training centres.

An account is given of various psychometric theories of stability and neurotic tendency among defectives, together with investigations into the prediction of occupational success; the chief finding was that tests alone are not valid predictors and that economic and social factors play an important part in success or failure, as would be expected. The section on occupational adaption in the community represents a beginning in a relatively unexplored field, but the findings must be regarded as tentative since the follow-up period was short. Certainly future work will have to concentrate more on this since most high-grade defectives are to-day dealt with not primarily because of occupational problems but on account of other social difficulties. The status of the imbecile in the community also requires further elucidation.

Part of the book deals with experimental studies of occupational training, education and psychotherapy with unstable patients, all of which yielded hopeful results. The authors conclude with a chapter on administrative and legal implications and offer some stimulating suggestions for changes in our methods of dealing with the mentally deficient. This section is somewhat controversial, particularly the argument that a special Mental Deficiency Act is no longer needed and that those at present certified could be dealt with by other existing services.

Partly because of the change in national socio-economic background, and partly stimulated by the findings of the M.R.C. Unit and other research workers, the approach to the problem of the mentally deficient is changing, and the custodial attitude is gradually being replaced by a more remedial one, or at least one which concentrates its efforts on capitalizing on the assets of the subnormal and making the best use of limited potentialities. Drs O'Connor and Tizard have demonstrated that this is a field in which social scientists can play an important and rewarding role. In a world where psychologists, psychotherapists, physicians and educationists have been and are only too prone to write-off the subnormal as uninteresting diagnostically and unrewarding prognostically, this book provides a refreshingly novel approach to the problem, and also a warning to those who are content to generalize on the basis of stereotypes which arose without adequate experimental observation in different socio-economic circumstances from those pertaining today.

A. D. B. CLARKE

Tredgold's Textbook of Mental Deficiency. By R. F. TREDGOLD and K. SODDY. London: Baillière, Tindall and Cox. Pp. xvi+480. 40s.

In this ninth edition of his father's original text, Dr R. F. Tredgold has succeeded in carrying on a notable tradition. With Dr Kenneth Soddy's help he has presented the point of view so well known since the first edition, and found also in the late Dr A. F. Tredgold's other writings. In certain ways, therefore, apart from modernizing detail (e.g. in the chapters on Pathology), the authors have not attempted to reorganize the available knowledge as an entirely new work.

The character and presentation of material differs from the previous (1952) edition in its treatment of psychology, and the social and administrative problem of mental deficiency. Some chapters, such as that on the Normal Mind, do not present anything surprisingly novel as compared with previous versions. Discussions of Conation, Affect and Cognition follow previous trends relying on Spearman and McDougall. However, in the next chapter, 'The Defective Mind', an entirely new approach to the problem of instability in defectives is suggested by one short section entitled 'Damaged' Children. This section reports some of the early work of Spitz and Bowlby and the work of Roudinesco from Dr Soddy's recent book on Mental Health and Infant Development. In a broad and general way, an explanation of 'moral deficiency' as defined in the Mental Deficiency Acts is offered in these terms, but the more recent work of Bowlby, and Bourne's *Lancet* (1955) article are not mentioned. No criticisms of the hypothesis of emotional deprivation are discussed.

Dr Tredgold and Dr Soddy examine the Sociology of Mental Defect in a chapter the latter part of which has been revised. The revision does not mean a change of approach, and the stress is still on the problem of the presumed increased incidence of mental defectives. The opinions expressed by Dr A. F. Tredgold in the columns of the *Eugenics Review* for 1908 are found again in this chapter, namely, fear of the increase of mental deficiency and fear of resulting national degeneracy. The authors do not expect that euthanasia, sterilization or segregation will help to solve the problem, and they devote their last chapter to a limited survey of methods of treatment which may give hope for the alleviation of the symptoms of mental defect.

This text needs to be assessed from the medical as well as the psychological point of view to do it justice. The authors disclaim any intention of offering an adequate description of the psychology of mental deficiency, and do not present any results of current psychological research. Its use to psychologists will therefore continue to be as a compendium of clinical medicine. In reading it, they must be careful to distinguish clinical impression from experimentally tested knowledge. In the field of mental deficiency, because of the lack of research, the former is often presented as the latter.

N. O'CONNOR

Studies in Feeling and Desire. By J. C. FLUGEL. London: Duckworth. 1955. Pp. xv + 200. 16*s*.

In its 'collected papers' quality, this book resembles *Men and their Motives* (1934) rather than *Man, Morals and Society* (1945). Yet, like all the late Prof. J. C. Flugel's major works, there is a central unity; in this instance it is one which will be familiar to those of us who had the good fortune to attend his courses on orectic psychology. It is 'feeling' (in the sense of pleasure—unpleasure) rather than 'emotion' which forms the central theme of these essays. The author is never happier than when dealing with the major, complex and subtle problems in this area; he examines their psychology with a surer touch than any other authority with whom the reviewer is acquainted.

The five chapters comprise: I, Sublimitation, its Nature and Conditions; II, Feeling and the Hormic Theory; III, *L'Appétit vient en Mangeant*; IV, The Death Instinct, Homeostasis and Allied Concepts; V, A Quantitative Study of Feeling and Emotion in Everyday Life. The first four of these papers are mainly theoretical and Chapter IV in particular has been considerably rewritten since its original publication. In Chapter I the author deals with sublimation in relation to 'more familiar' concepts like association, learning and conditioning. The succeeding two chapters, replete with stimulating thought and insights, lead up to the treatment of the death instinct; this is considered in terms of the idea of such varied authorities as Fechner, Claparède, Zeigarnik, Cannon (homeostasis), Wiener (Cybernetics) and Zipf (principle of Least Effort). In this fourth chapter and in Chapter III—the two which the reviewer found the most fascinating in a stimulating book—Flugel presents his own development of ideas deriving from James Drever (Sen.) concerning tension-reduction in its relation to the self-sustaining tendencies of motivation.

In this book we meet some of Flugel's most characteristic ideas and insights: his concern with formulation of laws of feeling; the principle of increase of satisfaction through inhibition (prominent in *The Psychology of Clothes* (1930) and *Men and their Motives*); his critical approach to Freud's concept of the death instinct, reformulated in terms of Cannon's homeostasis; and his concern to integrate the researches of Wolgemuth and numerous other investigators of emotion and feeling, who worked under his direction. Characteristic of this book is its concern with definition and greater methodological rigour of the kind which Flugel, like Ernst Kris and a few others, have sought to bring to psychoanalytic thinking. Ability to speak the languages of both what Cronbach (*Psychometrika*, 1954) has recently called 'Clinicia' and of what he has called 'Psychometrica' contributes to the strength of this book; its author does more than merely speak both languages fluently; he writes with the authority of being himself an experimentalist of stature together with the scholarship of an historian of psychology.

The originality, breadth and depth of Flugel's thought recommends this book; because of the richness of its insights and the reflection of its author's humanity, scholarship and wisdom, few psychologists can afford *not* to read it. The appearance of this book marks the passing of a major psychologist, and reminds us that its author, John Carl Flugel, has for many reasons, a secure place in the history of the science.

P. MCKELLAR

Developments in the Rorschach Technique. Vol. II. *Fields of Application.* By BRUNO KLOPFER and sixteen contributing authors. London: Harrap. 1956. Pp. xx + 828. 50*s*.

Because of the high level of one's expectation, it was perhaps almost inevitable that vol. II of 'the new Klopfer' should prove something of a disappointment. Unlike its predecessor (reviewed in this *Journal*, 1955, 46, 72–3) the present volume follows very closely the current fashion for publication in symposium form. Consequently, while there is a quite remarkable unity of theoretical orientation, the diversity of treatment is at times disconcerting, particularly in the case of contributions which bear signs, as some do, of having been prepared for use in some other context. Further, not all the contributing authors write equally well, and the reader who is sensitive to language may find himself wincing from time to time.

In his Foreword, Bruno Klopfer concedes the point regarding diversity; he also admits that he and his co-authors are 'just beginning to pile up convincing evidence for the statements made'. Edwin Shneidman, again, in his chapter 'Some relationships between the Rorschach technique and other psychodiagnostic tests', opens with the remark that 'the exact relationships among psychological tests are almost entirely a matter for future research'. The cynic will no doubt say that the truth of both statements is all too evident, and suggest that the present publication might well have been postponed. If one looks to the book for 'all the answers' one may be tempted to agree with our hypothetical critic. Nevertheless, its four parts, each devoted to a broadly defined 'field of application' (Genetic Psychology; Medical Psychology; Social Psychology, Anthropology, and Industrial Psychology; Diagnostic Practice and Projective Theory) contain much that is of value, not only in terms of Rorschach practice, but also in relation to the general study of personality in its varying aspects. Particularly interesting, from both points of view, are Gertrude Baker's chapters on 'Diagnosis of organic brain damage in the adult', and Gertrude Mieli-Dworetzki's on 'The development of perception in the Rorschach'. This latter is essentially a research report with normative data, which, while simply presented and conservatively interpreted, are adequate and convincing.

A third disclaimer, amounting almost to a main theme of the book, is summarized (by Hedda Bolgar) in the concluding sentence of the last chapter: '...we do not know too clearly what the boundaries are between the stimulus effect of the test material and the stimulus effect of the examiner....' This recognition of the test situation as a social or even therapeutic relationship represents a considerable advance in projective practice, and forestalls a good deal of criticism impugning dubious reliability and validity. In this connexion the reader is referred to Shneidman's chapter, already mentioned. Contrary to beliefs entertained in some quarters, Rorschach practice does seem to be moving in the direction of behavioural analysis and away from arbitary systems of interpretation from 'signs'.

The book ends with an elaborately classified Bibliography extending to over 100 pages, together with 'Cumulative Indexes' covering both volumes.

B. SEMEONOFF

Hearing Defects in School Children Scottish Council for Research in Education. 1956. Pp. x+112. 10*s*. 6*d*.

Environment and the Deaf Child. By STEVEN GETZ. Illinois: Thomas; Oxford: Blackwell. 1953. Pp. xvi+173. 27*s*. 6*d*.

In its latest publication, the Scottish Council is concerned with two problems: (1) the incidence of hearing defects among school children, together with their distribution and classification; and (2) the educational implications of these hearing defects. A complete year group (12-year-olds), in all 4170 children, were examined. The defects of hearing were ascertained by means of a pure-tone audiometer test carried out in two stages. Every child in the sample took part in a sweep test. Each child was tested at the 10 db. intensity level on the four frequencies of 250, 500, 1000, 2000 c.p.s. All those who failed at the 10 db. level on any one of these frequencies were given the full pure-tone threshold test of eight frequencies through 125 to 8000 c.p.s. In determining the relation between hearing defects and educational attainment, the results of the transfer examination test were used (Intelligence, English and Arithmetic), since results in these tests existed for nearly every child in the group. The investigation revealed many points of interest, and these are all clearly laid out in this book. Quite apart from its main aims and findings, this book is of special value for the many incidental but important points which are discussed; for example, the problem of differences in calibration among pure-tone audiometers, the question of unreliability among testers even with such an objective instrument as an audiometer, differences revealed in hearing loss on a re-test, after several weeks, the factor of noise in the examination of children in school, and the awareness or otherwise of teachers of hearing defects among their pupils. The report concludes with twelve recommendations, for each of which the investigation itself provides positive evidence. This book can be strongly recommended to educational and medical authorities who are planning pieces of work in their own areas.

The author of the second book is a psychologist in a school for the deaf in California. Getz took a group of 'relatively well-adjusted students and then a group of relatively maladjusted'*. These two groups were compared for each of the variables he proposed to study, his aim being to note any differences between the two groups on each of these variables. The variables were age at entering school, ability to communicate with the hearing and with other deaf pupils, educational achievement, social maturity, reading achievement, visual functioning and intellectual functioning. His conclusions for each item are briefly noted and discussed, and, in perhaps the most valuable part of the book, he discusses the arguments for and against the more controversial points held by other workers in the deaf field. The book is a careful but rather condensed study of the subject, and is suggestive rather than conclusive. Its 30 pages of bibliography will be of value to students in this field, and the book can be recommended as a useful synopsis of certain current problems in the education and upbringing of deaf children.

---

* Assessing adjustment by the Haggerty–Olson–Wickman Behaviour Rating Schedule A, and the Brunschwig Personal Adjustment Inventory for Deaf Children.

The Scottish Council's book scores by taking a relatively limited problem and treating it thoroughly. Getz, on the other hand, seems to be less convincing, perhaps because of the extent of ground he covers; but his book is full of ideas, and it certainly reveals the liveliness and sharpness of the controversial issues which circulate in the world of deaf education today.

H. H. HOOD

Non-parametric Statistics for the Behavioural Sciences. By SIDNEY SIEGEL. London and New York: McGraw-Hill. 1956. Pp. xvii + 312. 49*s*.

The publication of this, the first comprehensive text dealing with non-parametric statistical methods, comes most opportunely at a time when the forthcomong Fifteenth International Congress of Psychology announces 'Scaling methods and non-parametric statistics' as one of its symposium themes. Dr Siegel's book, however, is more than a compendium of statistical techniques; it provides an exceptionally clear and stimulating account of the aims and methods common to all statistical treatment of psychological data.

In his brief discussion of the theory of measurement the author distinguishes between four 'levels' of measurement, to the two more primitive of which the term is not, perhaps, usually applied. Measurement on the nominal scale would more commonly be described as classification, and that on the ordinal scale as ranking. It is with these types of measurement that non-parametric statistics is almost entirely concerned. Reference to the useful summary table of available statistical tests, which is conveniently printed as end-papers, shows that when one reaches the level of the interval scale (i.e. measurement as it is usually understood, though not necessarily with an absolute zero) available non-parametric methods cease, for most purposes, to be specially appropriate, in so far as they fail to make use of all the information contained in the data. Thus, there is no non-parametric measure of correlation which will handle data provided at the interval level; consequently, if the nature of one's population distribution does not conform to the requirement of normality assumed in the use of product-moment correlation, one has to choose between forcing the data into form by adjusting one's scale interval or losing information (and 'power') by falling back on a non-parametric test based on ordinal measurements—alternatives neither of which will appeal to many psychologists. Since the orientation of the present work is almost entirely towards hypothesis testing as the primary function of statistical investigation, the drawback is not so serious after all, at any rate in this context. However, the researcher who has to handle interval-scale measurements which conform to unusual distributions (e. g. J-shaped) will be disappointed—having been misled, perhaps, by the term 'distribution-free', which Dr Siegel wisely avoids—to find that the range of problems for which solutions are offered is very limited. Furthermore, one's eye having been caught by the recurrent heading 'Example for large samples', one learns with dismay that to the non-parametric statistician a 'large' sample is one in the neighbourhood of 20.

All this is intended not to imply criticism of Dr Siegel's excellent book, but to call attention to the as yet limited scope of non-parametric methods. It might indeed be preferable to make freer use of the term 'ranking' methods, as Kendall does—although this would not cover such techniques as the Walsh test for equivalence of related samples.

The many tables necessary for the application of the methods discussed are all contained in the Appendices, which extend to more than 50 pages. To have collected these, many of which have hitherto been relatively inaccessible, is in itself a service of inestimable value.

B. SEMEONOFF

Mr Lyward's Answer. By MICHAEL BURN. London: Hamish Hamilton. 1956. Pp. 288. 21*s*.

This is an account written by Michael Burn of the work carried out in the residential 'school' of Finchden Manor, Kent, a school for maladjusted adolescents, created by Mr Lyward and which is under his supervision and care. After fifteen years of teaching experience, the helping of difficult boys became his lifework.

About forty youths, ages 15–20, live together at Finchden at any one time, some sent privately by parents, others paid for by the local authorities. The atmosphere is one of comparative freedom and little restraint in which the boys find respite from the various pressures of life, and gradually return to normality. Most of the boys have been under the care of different types of specialists, and seem to be sent to Finchden as a last and desperate resort.

It is encouraging to find that, while some failures are reported, the majority of the boys who have passed through Finchden have made good in later life, and, in some cases, have had distinguished careers. The book contains descriptions of the boys and their difficulties, and one senses their struggles, and the nterplay of personalities in the incidents and conversations recorded. One also catches glimpses, now and again, of the seemingly casual treatment given by Mr Lyward, who gains the respect and affection of even the most hardened of boys. The details of one boy's problems and treatment make instructive reading.

The book is not systematic nor factual, and the first few chapters are disappointing in that they produce a somewhat confused picture of Finchden Manor and of its inhabitants. Some of the later chapters, however, are clearer cut and of greater value to the specialist, for, in these, an attempt is made to sort out some of the impressions of the writer and to assess the treatment. One cannot but admire the hard work of Mr Lyward and his staff, which is so carefully, if rather picturesquely, described by Mr Burn. The chapters as they unfold tell the same kind of story that other fairly similar places tell of children and youths with difficulties which they are unable to overcome by their own efforts. The book, as a whole, is of greater interest to the general reader than to the specialist. The latter, however, is happy to know that such excellent therapeutic work is being undertaken, and is appreciative of the manifold difficulties inherent in such a task.

MARY COLLINS

Unsettled Children and their Families. By D. H. STOTT. London: London University Press. 1956. Pp. 240. 16*s*.

This is a book which should have popular appeal for Probation Officers, Social Welfare Workers and Teachers. The type of case-work analysis suggests the delinquent child and his background.

The book is divided into two sections, the first being an expansion and enlargement of two lectures around which, the author tells us, the book ultimately grew. In this first section an examination is made of the effectiveness in practice of the social theory of humanitarianism, followed by criticisms of the present procedures for dealing with unsettled children. It is, however, questionable whether, in a book designed for a wide lay public, it is professional or suitable to present the author's criticisms of his own profession and their methods. One feels he is being less than fair to his colleagues. He stresses the need to deal with the family as a unit before we can deal with the individual child, and the importance of the study of relationships within the family. The science of family case-work is then discussed and an outline is given of some of the causes of breakdown in family life, for the guidance of Social Workers and others. He shows unsettled behaviour as a series of typical patterns, each representing stages of maladjustment. This is a useful chapter and contains much commonsense and helpful advice. A further examination is made of relationships within the family group, and a thesis is presented that, basic in all the breakdown patterns, is the feeling of the child, that its membership of a family group is in jeopardy. Twelve standard ways are then presented in which a child's security can be threatened. Consideration is then given to the principles of helping parents to understand why things have gone wrong. In his enthusiasm, however, to prove that emotional and environmental factors underlie all motivation and complications of family difficulties Stott denies intelligence a place. This assertion might be dismissed by many psychologists with long experience of dealing with problem families. A further chapter gives examples of cases with their methods of diagnosis and treatment.

In Section II of the book, an examination of his standard family patterns is made, and he returns to the conclusion that the solution, in helping these children, lies in understanding that their feelings of deprivation exist because they have no reliable group to which they can feel themselves permanently attached. The plurality of causes of delinquent behaviour is stressed frequently. The wider social causes of family breakdown are examined and finally he presents a theory of maladjustment. In this theory, which is reminiscent of McDougallian Instincts, the key to understanding human behaviour lies in the 'executive reactions' which to a certain extent he equates with instinctive reaction.

Family History Charts are given at the end of the book, and in the Appendix syndromes of parent-child attitudes which are useful for reference.

This is a book in which the author, although he says little that is new, has made a courageous attempt to classify, so far as possible, some of the environmental situations which lead to juvenile breakdown in behaviour and delinquency. This is a very difficult task, by reason of the width of the field; but if environment is not only exogenous but endogenous where do the claims of heredity come in? There is also a danger that the findings of psychology can be considered too simple and too direct. There are many controversial statements in the book, but it contains much commonsense and presents a realistic approach in trying to solve very complicated problems. J. C. MACLEAN

Personnel and Industrial Psychology. Second edition. By EDWIN E. GHISELLI and CLARENCE W. BROWN. New York: McGraw-Hill. 1955. Pp. ix + 492. $6.00.

Those who know the 1948 edition will need to be told only that the book has been completely rewritten although the layout and chapter headings are similar; that there are completely new chapters on selection and classification and on social factors in industry; that more than a quarter of the references are dated 1948 and later.

Those who have not read the earlier edition will want to know a little more. To my mind this 1955 edition is the best text-book on industrial psychology, for several reasons. First, it is well written, all topics are clearly set out, classified and defined. Secondly, it is scientific in temper, general statements and advice strictly following on research findings, and claims and generalizations of others are always submitted to critical assessment; logic is shown to be still a powerful tool. Thirdly, wisdom and sense act as cement between the separate sections of the book. Fourthly, it is up-to-date but with no trace of finality about it; no impression is left on the reader that a third edition will be unnecessary.

Examples of the authors' manner and matter are as follows. 'Discussions of fatigue are likely to be characterized by circular reasoning which completely invalidates any conclusions that are drawn.' 'It is not regarded as strange that a man will be motivated to avoid social contacts at breakfast time, while in the evening he actively seeks such contacts.'

The new Ghiselli and Brown, then, is *the* book for students who need a good survey of what personnel and industrial psychology is all about. But it is more than that: because of its scientific basis, it could serve as an interesting introduction to general psychology. It is short on theory, but I could imagine a head of a psychology department finding parts of this book as starting points, check points or illustrations to a study of theory in (say) social psychology.

Such criticisms as this reviewer can make fall into two groups. One group were found, on examination, to be not criticisms of Ghiselli and Brown at all but of the present state of psychological research. For example, the chapters on interviewing and on training in industry are rather thin. This is not the fault of the authors but simply a reflexion of the state of knowledge on these subjects. The second group of criticisms can probably be ascribed to patriotism: there are plenty of quotations from early I.H.R.B. and M.R.C. work, but not, perhaps, enough attention paid to more recent British research. However, this is rather an impertinent criticism of authors who cite nearly 400 references. DENIS MCMAHON

The Psychology of Occupations. By ANNE ROE. New York: Wiley; London: Chapman and Hall. 1956. Pp. xii + 340. 54*s.*

The home-made blunderbuss, schizilly fired from behind a pillar in the cloister, has become quite a thing in recent years. We have discovered that, provided we are careful to shoot a reasonable number of victims—fairly well mixed, perhaps, but with no really troublesome nonsense about sampling—a couple of days of slightly chaotic field sport will probably yield sufficient good clean and cosy post-mortem fun to keep us and our students occupied for long enough. If we work busily—using appropriate methods, of course—out of the machines in our morgues will tinkle factors to suit all tastes, few or many, orthogonal or oblique. Nowadays nearly everybody knows that only what goes in can come out. All the same, a factor's a factor. If we don't happen to like the look of it when we see it, we can always give it another push and interpret it.

So many seem to think. Is Dr Roe among them? Maybe: for in this book, which—shorn of its opening irrelevancies, and its final exhortations, called 'implications'—is a bold but unsuccessful essay in the classification of occupations, she succumbs in a big way to the temptation to pigeon-hole them according to ideas suggested by interest-inventory enthusiasts who appear to have got many of their initial questions wrong. True, she starts off with a conducted tour of some social anthropology; but this turns out to be a little pointless. Then she brings in a lot about Maslow and his basic needs, perhaps because she herself is greatly seized by one of them, the need for understanding, and is commendably keen to have a theory.

However, this too is a flop, and the whole boiling is quietly thrown overboard before the discussion on classification begins. Poor Maslow.

What Dr Roe is really out to do is to introduce us to an old friend in some new clothes. She categorizes occupations according to their 'group' and their 'level'. Groups of occupations are differentiated by reference to their 'primary *foci* of activity' (outdoor, business contacts, technology, service, and so on). The level of an occupation is determined by the skills it demands and the responsibilities it involves. In short, Dr Roe's scheme is built on the same basis as the one widely used in the Youth Employment Service in this country.

But it is not a very satisfactory basis. We may hope that Dr Roe will soon become as tired of it as we are; that she will soon see that most occupations are far too mixed to allow of their being regarded as 'belonging' to this group or that group; and that she will shortly start thirsting for a classification based on something nearer real life than her question-begging primary *foci*. It must be said, however, that of its kind her scheme is good; it is better thought out than ours; and what she has to say about it will certainly have to be taken into account by classifiers of occupations until somebody comes along with something fundamentally better. Her documentation is almost first-rate.

ALEC RODGER

The Psychology of Adjustment. (Second edition.). By L. F. SHAFFER and E. J. SHOBEN, JR. Boston: Houghton Mifflin; London: Constable. 1956. Pp. xxi + 672. 42*s*.

For this second edition of his 1936 text, now extensively revised and rewritten, Shaffer has as his co-author E. J. Shoben. The book begins with a very good introduction to motivation, learning and conflict. This is possibly the best section of the book, written with an economy and clarity of style that serves as a welcome reminder that difficult ideas need not be couched in difficult prose. There follows a discussion of the major adjustive mechanisms; adjustment by defence, by escape, by illness and so on. The next section describes the way in which organic and cultural factors determine personality, and the techniques used to describe and measure personality. This very useful section includes a chapter on Psychoanalysis. The book continues by describing how child-guidance clinics function, how psychotherapy works, and concludes with a discussion of mental hygiene in the school, at work, and in the home.

It would be remarkable if so ambitious a book were above criticism and, quibbles apart, three points seem important enough to deserve mention. The authors deal in a quite satisfactory way with the psychoneuroses, but the psychoses, organic brain lesions, the attack disorders and the toxic states meet with scant consideration. If these topics are outside the authors' terms of reference this should have been made clear. However, a text on the psychology of adjustment, howsoever defined, should grapple with the problem of depression and here again only a casual reference or two to the subject is included. Yet the psychiatric clinics seem full of people who are depressed because they have failed to adjust, or have failed to adjust because they are depressed. Finally, their several references to treatment leave one with the impression that if the patient is neurotic the treatment will be psychotherapy, unless the illness is a psychosomatic one, when recourse to drugs or surgery is permissible. This over-simplification may accord well with the authors' psychological analysis of maladjustment, but it leaves us with a distorted view of the work of the ordinary psychiatric clinic and mental hospital.

It will be seen that these criticisms refer to putative omissions. What has been said has been well said and well illustrated, and should provide an excellent introduction to the medical and non-medical student of psychology who intends specializing in this field. The references are fairly recent and a gratifying number of these cite British research.

JAMES DREWERY

The Social Psychology of George Herbert Mead. By ANSELM STRAUSS (Ed.). University of Chicago Press; Cambridge University Press. 1956. Pp. xvi + 296. 12*s*. 6*d*.

When George Herbert Mead died in 1931 he left but few publications behind him. He did, however, leave papers and notes and from them a group of students and colleagues compiled three volumes: *Mind, Self and Society*, *Movements of Thought in the Nineteenth Century*, and *The Philosophy of the Act*. In this book, Strauss has brought together selections from all three. Running through them all is the conception of the Self as socially derived through the meaningful gesture, which induces its performer to 'take on the attitude of the other'. This achievement of self-consciousness comes at the end of a long process of evolution, and Mead discusses the significance in philosophy of the evolutionary hypothesis. In the beginning was the Act, and in action 'the living being determines its environment'. 'Mental processes are just those phases of conduct into which the self as an object has come to deepen and render significant our analysis and to make possible the rational solution of our problems.' Mead as pragmatist is represented in the sections on 'The Power of Mind' and 'Stages in the Act'. He was naturally interested in the nature of scientific thought, and when he deals with history as a science he equally naturally takes the line that all history is contemporaneous—the past is the setting for the present and changes with it. Two sections are devoted to these matters.

These posthumous publications carry with them the defects of their origins. Since they were derived mainly from lecture notes, they are frequently obscure, sometimes incantatory, and all too often repetitious. These strictures apply least to *Mind, Self and Society* which is much the most coherent, and, indeed, a delight to read. The selections from that book have been checked, and they have been skilfully made. The other two books were not available.

One complaint must be voiced. No references to the original volumes is made; we do not know from what page or chapter the extracts have been taken. Perhaps Dr Strauss has not incorporated the 'generalized other' of the world of scholarship, but to those who have the ommission is exasperating. However, Mead is too little known in this country, and we must be grateful to Dr Strauss for making his work available in a compact form.

W. J. H. SPROTT

Mental Health and Mental Disorder. By Arnold M. Rose. London: Routledge and Kegan Paul. 1956. Pp. xii+626. 40*s*.

In this book, Rose has collected thirty-eight papers from forty-six contributors. On the whole, it is an admirable piece of work and the editor and eight subeditors deserve our thanks for what is, undoubtedly, the best available reference book on the sociological aspects of mental disorder. A quarter of the selections were written especially for this volume, and the rest with some exceptions are reprinted as they appeared originally. Consequently, there is some overlapping and repetition. For example, the Chicago studies of Faris and Dunham are referred to whenever the term 'ecological' occurs. These studies show that the incidence of mental disorder is highest at the centre of the city and declines in all direction towards the periphery. Since this finding is supported by similar studies carried out in nine other cities, it is a major point in favour of examining environmental factors which contribute to mental disorder.

This is not the book to refer to for any evidence of hereditary factors in mental disorder—as the subtitle says, it is 'A Sociological Approach'. In fact the word 'heredity' seldom appears, and although Kallman does get into a footnote, his name is not included in the index. Nevertheless, several contributors do mention the possibility of constitutional predisposition to mental disorder, even though this is not the main purpose of their papers. Otherwise the book is remarkably free of any doctrinaire bias. Some of the early chapters support a Freudian viewpoint, but Sewell cautiously presents evidence that would lead us to question the hypothesis that infant training affects childhood personality in the way predicted by psychoanalysts. Green is frankly anti-Freudian or at least hostile to the Horney-Fromm deviation, but his approach is hardly more than anecdotal.

Many different theories as to the causation of mental disorder will be found in this book. It is due to the failure of society to provide social roles, and the low degree of social organization in certain areas (Burgess); neurosis is the result of a conflict of roles due to the conflicting socio-cultural environment—a lack of integration between the primary and secondary selves (Dai). Primitive societies are relatively free from mental disorder because they are static, whereas American society is rapidly changing (Burgess); mental disorder is prevalent in primitive societies (Demerath). According to Burgess, neurosis arises because modern society is too complex—there are 50,000 different occupations to choose from—a view which is later contested by Schermerhorn. In America, it is flexibility and social change which are the principal sources of insecurity, while in Europe stratification and rigidity result in frustration (Frumkin, quoting Ruesch). People must be helped to adjust to the modern world (Burgess) but 'normality' is a middle-class ethic disguised as science (Schermerhorn). The debate continues—but each new speaker has not heard the previous speeches.

In this book there is plenty of well-intentioned preaching which will stimulate our sympathy but not our scientific curiosity; there is wild speculation, woolly thinking and dubious theorizing. But when these pages are eliminated, we are left with a rather smaller book which would still be well worth the price, and a constant source of data and references. Not all the papers are of equal value to the reader, but the 'Summary of Studies on the Incidence of Mental Disorders' (Rose and Stub) with its wealth of thought-provoking data would surely be counted as one of the ten essential contributions.

To summarize, the main purpose of the book is to challenge any complacency regarding our social responsibility for mental disorder. To this end the book contains a theoretical discussion of Social Psychiatry—emphasizing the contribution of psychoanalysis to sociology and, less emphatically, claiming a sociological contribution in return; a detailed summary of various studies analysing the correlates of mental disorder in terms of age, sex and such social factors as economic status, occupation, residence, marital status and race—it is interesting to notice how consistently there are higher rates for foreign-born as compared with 'native-born' Americans; a section on mental disorder in various communities—primitive, religious (the Hutterites), rural, semi-rural and cities; a sociological analysis of schizophrenia, psychopathy, drug and alcohol addiction; and some rather speculative papers on preventive and therapeutic measures. The challenge will not meet with the same response from all readers, but the book is invaluable to all who research or teach in the field of mental health and mental disorder.

GODFREY GARDNER

When Prophecy Fails. By LEON FESTINGER, W. RIECKEN and S. SCHACHTER. Minnesota University Press; London: Oxford University Press. 1956. Pp. vii+256. 32*s.*

It is conceivable that this book may usher in a new genre—the mild psychological thriller written by and for professional psychologists. The bulk of it is an exciting, often astounding, narrative about a contemporary group of Americans who predicted the coming of a second world-cleansing flood on a specified date, and believed themselves to be an élite who would be picked up by flying saucers (on which numbered seats would be reserved for them) and taken to a benevolent planet where they would be trained to return to and repopulate the earth. The authors, all members at the time of the Laboratory for Research in Social Relations of the University of Minnesota, displayed journalistic opportunism in insinuating incognito participant observers into the group shortly after the public announcement of the prediction and in maintaining an extremely thorough coverage by these rather 'cloak and dagger' means till the group dispersed after the failure of the prophecy.

Considered simply as an assiduously reported 'human interest' story, the book could hardly fail to appeal to anybody with normal curiosity in the wilder extremities of human belief and the resulting oddities of behaviour. But it was of course all done with a serious scientific purpose, namely, to show how and why it is that believers in a categorical prediction can emerge from its invalidation not only undaunted but with even stronger conviction and greater missionary zeal. The result is a sandwich-type book, the narrative forming the generous portion of meat between thin slices of scientific bread. One slice is an introductory chapter briefly surveying the history of millennial and Messianic movements and stating the hypotheses which were to be tested by direct observation of this particular movement. The other slice is a disarmingly frank methodological appendix which is in many ways the book's most interesting feature.

But even when methodological doubts are set aside, it is difficult to feel convinced that the book's scientific objective has been completely achieved. The so-called 'hypotheses' seems to be inductive generalizations based largely on the historical survey of millennial movements, and they are stated so as to specify five conditions necessary for an increase in proselyting to follow an unfulfilled prophecy. The authors concede, however, that failure of a prophecy to come true is sometimes followed by the complete break-up of the movement, with no further attempts to convert others, even though the five conditions have been satisfied. They do not pursue this damaging admission, possibly because it would surely lead them into an analysis of the personality dynamics of individual members of the group, and they have throughout studiously avoided the orthodox 'genetic' analysis in favour of a Lewinian group dynamics approach. In consequence, it all reads like an odd case-history—odd because it concerns a group rather than an individual, and because practically all the stock motivational interpretation has been deliberately left out. Such a self-denying ordinance seems only to make the task needlessly difficult and perhaps impossible. At any rate, it is arguable that even a fictional version of the more usual motivational analysis (like Joyce Cary's account of the Adventists in his *Except the Lord*) brings more insight than this rather blinkered approach. Perhaps this is not entirely fair comment, for the authors finally restrict themselves to explaining 'why increased proselyting follows the disconfirmation of a prediction', and their explanation is 'one derivation from a general theory' of dissonance which is to be elaborated in a forthcoming book by Leon Festinger. Possibly, therefore, this book merely suffers in this respect from premature publication.

Notwithstanding these misgivings, this is undoubtedly a field study of remarkable novelty and unothodoxy which brings several important questions into sharp focus.

A. J. M. FLOOK

Psychology, Psychiatry and the Public Interest. Edited by MAURICE H. KROUT. Minnesota University Press, London: Oxford University Press. 1956. Pp. xv + 217. 32*s.*

The relationship between clinical psychology and psychiatry is the main topic of this book. Its aim, according to the editor's preface, is peace, 'based on mutual understanding and mutual recognition' which 'can come only from a broad and free discussion of the relationship between psychiatry and psychology by qualified representatives of both professions'. This being so, it is surprising to find that, of the fifteen individual contributors, only one (Paul E. Huston) is a psychiatrist. The remaining contributors are psychologists in the U.S.A., most of them working in a medical setting or in private practice. Eight of the chapters have already appeared elsewhere. This uneven balance of forces and the tone of some of the chapters tend to belie the preface, and a good half of the book might well have borne the title 'The Case for the Psychologist in Private Practice'. This does not mean that it has no relevance for British psychology. In the course of presenting the case, opinions are expressed on matters of great importance for both professions here and elsewhere. Opinions, however, do not win arguments and the supporting evidence is scanty.

A question on which there is much agreement between the psychologists and complete disagreement between them and the medical attitude as expressed in the book is, in the title of one chapter, 'Who Shall Practice Psychotherapy?' Many of the psychologists argue that the treatment of neurosis can be satisfactorily handled by an appropriately trained psychologist working without medical supervision. Robert M. Lindner, to whose memory the book is dedicated, goes so far as to suggest, in a posthumous reprint of an earlier paper, that present-day medical and psychiatric training does not produce practitioners competent to deal with psychopathological cases. 'Among the medical practitioners of psychotherapy of this generation', he writes, 'the majority are but spottily educated and informed. In the place of a broad systematized knowledge of man, his culture, and his works this majority possesses a handful of clinical facts wrapped about with terminological vagaries and ill-considered generalizations.' One can feel much sympathy for this point of view and yet have an uneasy feeling that its proponents are themselves speaking from insecure positions. One looks in vain for evidence that either group is better than the other when it comes to helping the people in need.

A point which is made several times throughout the book, and which is used to justify the above position, is that neuroses and character disorders are *not* medical problems and that psychotherapy is *not* a medical procedure. The first chapter, by Nevitt Sanford, makes some interesting observations in this connexion, and the point is developed by several contributors, particularly Carl R. Rogers. The idea that many psychiatric patients are not really ill but are seeking help for the solution of problems of living which they feel unable to solve unaided must often strike those who work with neurotics. This is, perhaps, the central problem posed by the book. It is an important problem and one which deserves a much more thorough examination than it is given by Dr Krout and his collaborators.

The authors have earned our gratitude for bringing into the open a number of important and controversial issues. It is to be hoped that somebody will now be stimulated to undertake the more difficult task of examining these issues more thoroughly in the light of all the evidence he can obtain.

J. G. INGHAM

Das autogene Training. (*Konzentrative Selbstentspannung.* Ninth edition). By J. H. SCHULTZ. Stuttgart: Thieme, 1956. Pp. xx + 351. D.M. 29.

Autogenes Training is understood as a systematic, concentrative relaxation without any specific suggestion. The aim is to achieve a general switching of the normal state of wakefulness to a generalized feeling state of low awareness. This frees the subject from all his specific problems and induces states of warmth, heaviness and relaxant feelings. A wide variety of accompanying physical and psychological concomitants, as well as the relationship to a number of related techniques, are discussed in detail. The method is, in individual or group administration, applied to a variety of disturbances of the vasomotor, respiratory, digestive or central nervous systems. Thus, non-neurotic patients, psychoneurotics (with the exception of deep-rooted chronic states), and a variety of psychopathic personalities are said to benefit distinctly to a degree of 63%. If one compares the ninth with the first edition (1932), it becomes immediately clear that neither method nor theory has changed to any considerable extent. In fact, no new data of an experimental kind have been added. Any other recent literature that has been considered is of the practical clinical type. Anglo-American literature is only occasionally referred to, and no systematic reference is made to the great number of experimental psychological investigations that exist in this field. It would appear to the reviewer that such an attempt might have avoided a certain staleness and sterility in an otherwise important and probably rewarding field of clinical research.

J. C. BRENGELMANN

OTHER PUBLICATIONS RECEIVED

Hypnotherapy with Children. By G. Ambrose. London: Staples. 1956. Pp. 135. 10*s.* 6*d.*

Midwest and its Children. By R. G. Barker and H. F. Wright. Illinois: Row, Peterson. 1955. Pp. 532. $7.50.

Lists of Researches in Education and Educational Psychology. Supplement I. By A. M. Blackwell. London: Newnes. 1956. Pp. 62. 15*s.*

Social and Moral Attitudes. By S. Brahmachari. Calcutta: Orient Longmans. 1954. Pp. 113. Rs. 5.

The Functions of Social Conflict. By L. Coser. London: Routledge and Kegan Paul. 1956. Pp. 188. 18*s.*

Six Children. By E. J. Foote. Illinois: Thomas; Oxford: Blackwell. 1956. Pp. 317. 42*s.*

The Negro Potential. By E. Ginzberg. Columbia University Press; London: Cumberlege. 1956. Pp. 144. 24*s.*

Hypnotism and Crime. By H. E. Hammerschlag. Translated from German by J. Cohen. London: Rider. 1956. Pp. 148. 13*s.* 6*d.*

The Human Hand: the Living Symbol. By N. Jaquin. London: Rockliff. 1956. Pp. 170. 21*s.*

Sigmund Freud: Four Centenary Addresses. By E. Jones. London: Tavistock. 1956. Pp. 150. 18*s.*

The Idiom of Contemporary Thought. By C. Knox. London: Chapman and Hall. 1956. Pp. 206. 18*s.*

The Pollen Path: a Collection of Navajo Myths. By M. S. Link. Stanford University Press; London: Cumberlege. 1956. Pp. 210. 48*s.*

Psychology, Evolution and Sex. By C. P. Martin. Illinois: Thomas; Oxford: Blackwell. 1956. Pp. 166. 36*s.*

The Mind Alive. By H. and B. Overstreet. London: Davies. 1956. Pp. 288. 16*s.*

You and Psychosomatics: a Guide to Mind-body Health. By J. A. Winter. London: Rider. 1956. Pp. 172. 12*s.* 6*d.*

L'Analyse Factorielle et ses Applications: Colloques Internationaux du Centre National de la Recherche Scientifique, no. lxv. Paris. 1956. Pp. 431. 1500 F.frs.

The Problem of Knowledge. By A. J. Ayer. Harmondsworth: Penguin. 1956. Pp. 224. 3*s.* 6*d.*

Automation in Theory and Practice. By E. M. Hugh-Jones (Ed.). Oxford: Blackwell. 1956. Pp. 140. 12*s.* 6*d.*

Chisungi: a Girl's Initiation Ceremony among the Bemba of Northern Rhodesia. By Audrey I. Richards. London: Faber and Faber. 1956. Pp. 244. 42*s.*

Mental Health and Education in Hong Kong. By K. E. Priestley and B. R. Wright. Hong Kong University Press; London: Cumberlege. 1956. Pp. 97. 7*s.* 6*d.*

Conditioning and Sexual Behaviour in the Male Albino Rat. By K. Larsson. Acta Psychologica Gothoburgensia. Stockholm: Almqvist and Wiksell. 1956. Pp. 269. Sw.kr. 23.

Human Disease. By A. E. Clark-Kennedy. Harmondsworth: Penguin. 1957. Pp. 267. 3*s.* 6*d.*

Sex in Christianity and Psychoanalysis. By W. G. Cole. London: Allen and Unwin. 1956. Pp. 329. 21*s.*

Religious Factors in Mental Illness. By W. E. OATES. London: Allen and Unwin. 1957. Pp. 239. 16*s.*

The Image. By K. BOULDING. Michigan University Press; London: Oxford University Press. 1957. Pp. 175. 30*s.*

Juvenile Delinquency Proneness: a Study of the Kvaraceus Scale. By J. K. BALOGH and C. J. RUMAGE. Washington: Public Affairs Press. 1956. Pp. 35. $1.00

Human Relations in International Affairs. By S. W. BEARDSLEY and A. G. EDGELL. Washington: Public Affairs Press. 1956. Pp. 40. $1.00. (A selected bibliography of 117 titles.)

Edward Sapir: Culture, Language and Personality: Selected Essays. By D. G. MANDELBAUM (Ed.). California University Press; London: Cambridge University Press. 1956. Pp. 207. 13*s.* 6*d.*

Position and Subject-matter of Criminology: Inquiry Concerning Theoretical Criminology. By H. BIANCHI. Amsterdam: North-Holland. 1956. Pp. 216. Guilders 12.

In Their Early Twenties: a Study of Glasgow Youth. By T. FERGUSON and J. CUNNISON. London: Oxford University Press. 1956. Pp. 110. 12*s.* 6*d.*

Religion and the Psychology of Jung. By R. HOSTLE. Translated by G. R. LAMB. London: Sheed and Ward. 1957. Pp. 249. 16*s.*

Mind and Body. By P. L. ENTRALGO. London: Harvill. 1955. Pp. 150. 12*s.* 6*d.*

How to be Happy though Human. By W. BERAN WOLFE. Harmondsworth: Penguin. 1957. Pp. 368. 4*s.*

Effective Personnel Selection Procedures. By C. H. STONE and W. E. KENDALL. London: Staples. 1957. Pp. 433. 42*s.*

The Character of Man. By E. MOUNIER. Translated from French by C. ROWLAND. London: Rockliff. 1956. Pp. 341. 42*s.*

Understanding Minority Groups. By J. B. GITTLER (Ed.). New York: Wiley; London: Chapman and Hall. 1956. Pp. 139. 26*s.*

Psychologie der Sprache. By F. KAINZ. Stuggart: Enke. 1956. Pp. 537. DM. 49.80.

Personality in a Communal Society: an Analysis of the Mental Health of the Hutterites. By B. KAPLAN and T. F. A. PLAUT. Kansas: University Press (Social Science Studies). 1956. Pp. 116. Cloth $3.25; paper $2.50.

Toward Better Personal Adjustment. 2nd ed. By H. W. BERNARD. London: McGraw-Hill. 1957. Pp. 454. 41*s.* 6*d.*

Beyond Laughter: a Psychoanalytic Approach to Humor, Wit and the Comic. By M. GROTJAHN. London: McGraw-Hill. 1957. Pp. 285. 45*s.*

Burning Water: Thought and Religion in Ancient Mexico. By L. SÉJOURNÉ. London: Thames and Hudson. 1957. Pp. 192. 25*s.*

Freudiana: including Unpublished Letters from Freud, Havelock Ellis, Pavlov, Bernard Shaw, Romain Rolland, et alii. By A. A. ROBACK. Cambridge, Mass.: Sci-Art Publishers. 1957. Pp. 240. $5.00.

Family Mental Health and the State. Proceedings of the Eighth Annual Meeting of the World Federation for Mental Health, Istanbul, August 1955. London: Lewis. No date given. Pp. 165. 15*s.*

Juvenile Delinquents in a Psychiatric Clinic. By W. LITAUER. London: Institute for the Study and Treatment of Delinquency. 1957. Pp. 59. No price given.

Descartes's Rules for the Direction of the Mind. By H. H. JOACHIM. (Edited by E. E. HARRIS.) London: Allen and Unwin. 1957. Pp. 124. 10*s.* 6*d.*

VOLUME 48 AUGUST 1957 PART 3

THE DISTRIBUTION OF INTELLIGENCE

BY CYRIL BURT

Current criticisms of the use of the normal curve in assessing intelligence seem to arise from misconceptions both about the nature of the evidence available and about the hypothesis assumed. The theoretical evidence suggests that individual differences in intelligence are the effect partly of multifactorial and partly of unifactorial inheritance: on this basis the distribution to be expected would be bell-shaped, but more or less skewed, i.e. a 'type IV' rather than a 'normal' curve. The empirical evidence is mainly derived from tests scaled independently of the resulting distribution. With such methods, provided the samples are sufficiently large and complete, the frequency-curves obtained exhibit small but significant deviations from strict normality. If, however, specific cases of mental deficiency are omitted, then, for most practical purposes, the resulting distributions can be treated as approximately normal.

I. THE DEFINITION OF INTELLIGENCE

The question which Mr Lewis has examined in his paper on 'the normal distribution of intelligence' (Lewis, 1957) has been raised by several recent writers—particularly those who criticize what they call 'psychometry'.* Dr Heim (1954), for example, has argued that 'frequency distributions on any test depend mainly on the particular system of scoring adopted'; she therefore infers that the notion that intelligence is distributed normally can be nothing but a gratuitous assumption, though 'often hailed as a scientific discovery'. Mr Kenneth Richmond, in an earlier article in this *Journal* (1953), advances a similar contention: 'the psychometrist', he believes, can only ensure the '*a priori* principle' of normality by 'tinkering with the test material'; hence 'measurements appear normally distributed simply because the test has been so constructed that they shall be distributed in that way'. Mr Lewis apparently inclines toward much the same conclusion.

In considering such criticisms it is essential at the outset to make sure that, when they speak of intelligence or of its distribution, both the critics and the writers criticized are referring to the same thing. The word intelligence has such a wide variety of meanings that it is tempting to borrow Spearman's device, and substitute symbols for phrases. Here and elsewhere, when I use it in a technical sense, I intend it merely as a brief and convenient designation for the 'inborn general cognitive factor' (Burt, 1911, p. 95; cf. Hebb, 1949). This fourfold expression seems best to convey what Spencer, Galton and Binet had in mind whenever they employed the term. If, however, we adopt Fisher's convenient practice of using Greek letters for hypothetical variables and Roman for the empirical estimates, then, as I have elsewhere suggested (1955 *a*, p. 88), we might avoid much of the ambiguity by labelling this abstract and theoretical concept γ (gamma), in contradistinction to Spearman's g, which is commonly used to denote measurements derived from actual tests.

Nevertheless, each of the semi-popular terms in this familiar definition has at times occasioned some misunderstanding. Hence each calls for a word of explanation. (i) To those unacquainted with the jargon of factor theory, the term 'factor' is apt to suggest a concrete ability that can be identified by ordinary observation. Thus Dr Heim rightly observes that we cannot 'separate the cognitive factor' from other aspects of mental activity. Yet, though we cannot separate it, we can certainly distinguish it; and for this purpose the factorist employs a method of analysis which is virtually identical with the physicist's method of resolving a resultant force into its hypothetical components. His meaning therefore would be more clearly expressed if he retained Pearson's original term 'component' instead of re-naming it a 'factor'. (ii) The so-called 'general' factor is really a 'group factor', since it is assumed to enter into all measurements of 'cognitive' efficiency, but not into assessments of emotional or temperamental characteristics. As has often been remarked, a truly 'general' factor would embrace something much wider than intelligence, and might be influenced quite as much by motivation as by sheer capacity. (iii) The word 'cognitive' is here used in contradistinction to 'orectic'. But it is often mistakenly supposed to mean purely intellectual activities as contrasted with practical—knowing as distinct from doing. Accordingly, it would perhaps be less ambiguous to introduce yet another technical term, and speak of a 'cybernetic' as contrasted with a 'dynamic' component. (iv) Finally, the word 'inborn' is apt to convey popular and out-of-date notions of 'inheritance', and to suggest that 'hereditary' characteristics can be sharply separated from those that are 'acquired'. The objections urged by Mr Richmond and Dr Heim against this rather naïve conception every geneticist would nowadays endorse. But that of itself does not suffice (as Dr Heim apparently holds) to 'expose the absurdity of such statements as "the (estimated) contribution of innate factors is probably not far from 75 or 80 per cent."'. I myself use such words as 'innate' or 'inborn' merely as short and simple ways of designating 'that part of the variance of the general factor which is attributable to genetic constitution'.

---

* In this paper the word is used in the sense defined by Dr Heim (1954, pp. 1, 22) rather than by Guilford (1936, p. xi); but I do not agree with the sharp division she makes between psychometrists on the one hand and psychologists on the other. A psychometrist who was not first and foremost a psychologist would, in my opinion, be a very bad psychometrist.

It is, of course, open to the critic to deny that any such component exists, or, if it exists, that it plays any considerable role in determining human behaviour. Elsewhere* I have summarized the vast amount of evidence now available on these points: and, so far as I am aware, no critic has attempted a systematic refutation of the arguments.

There are, however, two alternative interpretations of the term intelligence that are still widely current.

(1) It is frequently urged that what is needed is not a hypothetical but an operational definition; and those who advocate this standpoint commonly adopt the formula that 'intelligence is the mental characteristic measured by intelligence tests' (e.g. Fraser, 1954, p. 11). Intelligence as thus defined I have called g_t: it corresponds fairly closely with Spearman's assessments of g, which were based, not on factor analysis, but on certain tests selected as 'reference values'. Apparently this is the definition on which Mr Lewis relies in the latter part of his paper. In practice, however, to define intelligence in this circular manner is as unhelpful and as uninformative as it would be to define temperature as 'the physical characteristic that thermometers measure'. To begin with, how are we to determine what can be accepted as tests of intelligence and what cannot? Mr Lewis seems to regard the Moray House tests as the best and most typical; others would take the Terman-Merrill tests as their standard; others again prefer tests of a non-verbal or performance type. Furthermore, all such tests have imperfect correlations

* The definitions given above have been more fully explained in earlier publications on the subject (1911, 1912, 1917, 1921) and the evidence reviewed afresh in several recent papers (1943, 1955, and refs.).

both with each other and with independent assessments. Thus, as with other forms of psychological measurement, their results are largely affected by incidental influences irrelevant to our main purpose, and are consequently disturbed, to a far greater extent than physical measurements, by error. But if we have no antecedent definition of what we want to measure, how can we distinguish what is irrelevant and what is not?

(2) The majority of the more recent critics, however, would like to get right away from the notion that intelligence is to be defined in technical or quantitative terms. Dr Heim, for instance, complains that the postulate adopted by the statistical psychologist—that 'there exists some one attribute, "intelligence", which is one-dimensional, measurable, capable of quantitative but not qualitative differences'—is far too remote from the 'ordinary meaning of the word as the layman understands it', and that 'to take a popular word, denude and disguise it and try to restrict its circulation', is 'quite indefensible'. Similarly, Mr Richmond describes the concept of 'measurable intelligence' as an *ignis fatuus* which leads the student into the region of statistical jargon where he 'soon finds himself bogged down in mathematical abstruseness'. The proper procedure, we are told, is to assign the word 'a meaning that is comfortably compatible with that of the layman'. Thus, when Dr Heim and Mr Richmond speak of 'intelligence', they treat it as a synonym for what I dubbed g_i (Burt, 1955*a*, p. 86).

In point of fact, however, 'intelligence' is not a word of popular origin. It was coined as a technical term to express a highly technical concept. Like other scientific or philosophical expressions, it has been borrowed by lay writers, and used so loosely as to be almost worthless for purposes of exact logical argument. But why should the layman dictate to the scientist how his terms are to be defined? No one would think of telling the chemist that his definition of a gas should be governed by the way the word is used in common parlance, or argue that, because the kinetic theory of gases can only be expressed in 'statistical jargon' and with 'mathematical abstruseness', it should be discarded in favour of something 'comfortably compatible' with the untrained layman's comprehension.*

But, whether we accept these linguistic criticisms or not, they do not of themselves rebut the psychometrist's contention that intelligence *as he defines it* is distributed in approximate conformity with the normal curve. Dr Heim and those who follow her seem to have mistaken a disagreement over names for a disagreement over facts. No psychometrist, I imagine, would dispute her statements that 'men differ qualitatively as well as quantitatively', or that they vary in the traits she calls 'social intelligence', 'creative intelligence', 'verbal intelligence', and the like: what he objects to is the practice of lumping together a number of diverse and complex characteristics—some involving highly specific abilities, others involving temperamental as well as cognitive components—under one and the same term 'intelligence'.

Mr Lewis's attack on the normal curve goes much deeper. He does not, if I understand him rightly, quarrel with the psychometrist's definition. On the contrary, accepting that definition (or something very near it), he adduces several arguments to show that the claim that intelligence is normally distributed (*a*) is 'by no means self-evident',

---

* Since the above was written Mr T. R. Miles has published the abstract of an admirable paper 'On Defining Intelligence' (*Brit. Psychol. Soc. Bull.* (1956), **29**, 2), in which he criticizes (amongst others) these last two definitions from the standpoint of modern logic. I should, I think, readily assent to all the objections he has advanced.

(*b*) has not been 'established from the results of intelligence tests', and (*c*) cannot be 'deduced from genetic grounds'. In the end he too comes to much the same conclusion as Dr Heim, who holds that the hypothesis of normal distribution is supported by no evidence or proof, and is therefore at best merely a convenient fiction and at worst an unwarrantable distortion of the original scores. The psychometrist postulates normality when standardizing his tests; and therefore the 'quotients' secured with such tests conform with the normal curve, but that is merely an artificial consequence of the assumption he has arbitrarily made.

Now I venture to think that all this gives the reader a rather unfair picture of the type of proof on which those who are criticized by Mr Lewis and Dr Heim have actually relied. Like other generalizations in natural science, our conclusions about the nature and distribution of intelligence rest on what Whewell has called 'a consilience of inductions'. They are supported, not by just one kind of evidence (such as 'the distribution of quotients'), but by several converging lines of argument. Being inductive, each inference by itself can lead to no more than a probable conclusion, and its probability, considered in isolation, may be relatively low; but the *cumulative* effect of the different arguments is overwhelming.

The effect of 'consilience' is so persistently overlooked that it is worth while emphasizing its importance. Let us suppose that (as argued below) there are five hypothetical curves that deserve empirical investigation, and that the antecedent probability (based on genetic theory, and the like) in favour of the second of these (a moderately asymmetrical curve of type IV) is twice as great as that for any other single type: this would evidently give it an *a priori* probability of 2 out of 6, i.e. 1 out of 3—a very modest estimate in view of the relative infrequency of the other curves in previous genetic work. And now let us imagine that, in general, the odds in favour of a verdict given by any one of the independent sources of empirical evidence described are 4 to 3—once again an extremely conservative estimate. There are three such sources. Accordingly, on applying the usual formula, it can readily be shown that, if all three sources of evidence agree in confirming the second hypothesis, the *a posteriori* probability that this hypothesis is correct will be 0·97: in other words, the odds in favour of that hypothesis have risen from 1 out of 3 to 31 out of 32 (for the method see Keynes, 1921, pp. 182f.). Of course, no one pretends that such probabilities can be exactly assessed, or that such formulae can be regarded as anything but simplified approximations: but they serve to bring home in a striking way the cumulative force of converging lines of evidence.

It is all too easy for a critic to seize on one particular argument and to ignore the fact that the reasoning is avowedly based on a comparison of probabilities: he can then demonstrate that this one argument is inconclusive, i.e. that it can claim the certainty neither of a 'discovery' nor of a 'self-evident proposition', nor yet of a formal 'deduction'. But plainly this is, in the most literal sense, an *ignoratio elenchi*. What the critic has to do is not simply to demonstrate that a probable conclusion cannot claim the certainty of a deduction or a self-evident postulate, but rather to show that *his* hypothesis has a higher probability than the one which he is criticizing; and this can best be done by adducing fresh empirical data from an appropriately planned inquiry of his own.

II. THE ALTERNATIVE HYPOTHESES

The object of our inquiry is to determine the distribution of *innate* ability. We can admit straight away one point repeatedly stressed by the critics—namely, that the results of every test, taken as they stand, are bound to be affected by environmental and other post-natal influences. What I myself have tried to show is that at least 75% of the total variance must still be due to genetic influences, and probably far more (1955*b*, p. 175 and refs.); and this should be sufficient for our purpose.

Dr Heim, however, tells us that her first aim is to expose the absurdity of all such statements (1954, p. 124). Her argument is that any attempt to estimate the relative contributions of 'nature' and 'nurture' 'must imply a separation which does not exist'. But her contention, if valid, would manifestly undermine almost every attempt at biometric analysis: for, as the geneticist constantly reminds us, '*every* character can—and in fact nearly always does—show heritable and non-inheritable variation simultaneously' (Darlington & Mather, 1949, p. 64). Her difficulties, I fancy, really arise from a confusion between the individual and the population. The analysis that she proposes to refute is really an analysis of the variance of a *population*. Such an analysis, she says, 'taken to its logical conclusion leads to such absurdities as the following'; and she then gives two imaginary analyses of an *individual's* deviation ('Tommy's') from the parental mean in terms of the contributions of heredity and environment respectively. But this is *not* a logical conclusion; it is a logical fallacy—a fallacy so common that Aristotle and his followers gave it a special name, παρὰ τὴν διαίρεσιν. It is like arguing that because 15% of the income of the *population* is unearned, therefore 15% of *Tommy's* income must be unearned.

Fisher has succeeded in deriving (on the basis of certain reasonable conditions) perfectly valid formulae for analysing the variance of a population; and in the case of stature and other physical characteristics has obtained plausible estimates for the relative contributions to that variance of (*a*) 'heredity' (defined as including 'genotypes', 'dominance', and 'homogamy') and (*b*) 'environment' ('using that term formally for causes independent of heredity') (Fisher, 1918, pp. 422–4). It is difficult to see why arguments and procedures used in the case of physical measurements cannot also be applied in the case of mental measurements. In both cases we have to allow for minor errors of measurement and for the effects of environment (defined as above); and in the case of mental measurements these two contributory factors will no doubt be somewhat larger. But I can discern no difference of principle.

Dr Heim and Mr Lewis would apparently condemn all such methods as invalid. Yet they never explain where precisely Fisher, and those who follow him, go wrong. And, since Fisher's general views are accepted by practically all geneticists, it is surely permissible, until a more direct refutation is offered, to accept them here.

However, although we may now agree that the same kinds of genetic mechanisms determine both physical and mental variations, the detailed processes involved may of course differ appreciably in the case of this particular characteristic or that. What then are the alternatives that we have to consider in the case of intelligence, and what types of distribution are they likely to produce? Within the framework of current genetical theories there are, I think, four distinct hypotheses that deserve to be examined: and each leads to its own distinctive form of frequency-curve.

(1) *Symmetrical and normal.* The simplest hypothesis would be to assume that differences in intelligence are due solely to multifactor inheritance, and to treat the effects of subsidiary processes—dominance, linkage, assortative mating, and the like—as too slight to affect the ultimate results. On this basis, as Karl Pearson showed (1903), we should expect a perfectly normal distribution. This simplified mechanism is the one which Mr Lewis apparently supposes that the psychometrist postulates. (In his paper, as printed, he now agrees that 'this is undoubtedly an oversimplified mechanism'.) But this rather primitive picture disregards a number of well-established facts, and leads to correlations between relatives far lower than those actually observed (Burt, 1955*b*, p. 174). Certain types of extreme deviation seem undoubtedly due to single genes; and in some types the responsible gene is dominant, in others recessive, and in others again sex-linked (Burt, 1935; Fraser Roberts, 1940). In the clearest cases the deviation appears to be subnormal and well marked; but there is evidence, far less conclusive, to suggest that occasionally the deviations so transmitted may be relatively slight, and in others perhaps supernormal. Moreover, the resemblance between parents due to assortative mating can hardly be less than 0·40. There can be little doubt, therefore, that the conditions to be postulated are not those adopted by Pearson, but the far more general conditions assumed by Fisher in his later inquiry into the mathematical consequences of the Mendelian theory (1918).

(2) *Moderately asymmetrical.* The hypothesis that I myself tentatively put forward (e.g. 1935, pp. 79f) was that, for intelligence as for height, weight, and many other graded characteristics, we seem compelled to postulate two modes of inheritance: (i) in certain cases, as the pedigrees indicate, the deviation studied may act like a recognizable trait dependent on a *single* gene—probably in most cases partially dominant, and sometimes sex-linked; (ii) in other cases it is apparently determined by the joint action of a *large number* of genes. The 'major genes' seem comparatively rare, but each will produce effects that are large and (it would appear) for the most part detrimental; the 'polygenes' must be much more numerous, but their effects will be too small to be identified individually. With this double assumption, the resulting distribution would take the form, not of the normal curve, but of an asymmetrical, bell-shaped curve of unlimited range in either direction. To borrow Pearson's convenient nomenclature (1914, p. lxiii), it would be 'type IV' rather than 'type G' (the 'normal' or Gaussian, which is in fact a limiting form to type IV). Thus, as a result of this twofold method of transmission, 'the inheritance of intelligence would resemble that of stature, though with a more marked divergence from strict normality' (Burt, 1935, p. 81).

(3) *Highly asymmetrical: negatively skewed.* The foregoing are by no means the only hypotheses that demand consideration. Several writers have insisted that genetically intelligence must be an exceedingly complex quality (a view which I should not for a moment deny—so indeed is stature); and from this they infer that 'an exceptionally high degree of intelligence must be the resultant of exceptionally large deviations, each exceptionally rare, affecting numerous special abilities'; these favourable conjunctions, they consider, must be 'decidedly infrequent'. Hence 'the number of persons superior to the mode must be much smaller than the number of those inferior'. Such a theory would produce a distribution with a pronounced degree of negative asymmetry, since the upper tail would be markedly compressed (cf. Snyder, 1940, ch. xiv).

(4) *Highly asymmetrical: positively skewed.* Finally, one or two writers have suggested

that the effects of the genes 'might combine by multiplication instead of by addition'. In that case differences in net ability would correspond more closely with differences in output, and so, indirectly, with differences in income. On these and other grounds, Pareto and his American followers have inferred that the distribution of innate ability must be highly asymmetrical and possess an 'elongated upper tail'. Thus, as Pareto himself observes, the highly skewed curves for the distribution of income are 'not a secondary consequence of *economic* conditions, but reflect a *biological* law' (cf. Davies, 1941, pp. 427 f.).

(5) *Other types*. Mr Richmond thinks that 'the true distribution of intelligence may well follow a curve of its own'. But neither he nor Dr Heim indicates what particular type of curve they have in mind. Mr Lewis, it would seem, is inclined to believe that its form would be approximately 'rectangular'. But he offers no evidence from genetics to show how this could occur.

Had we no empirical data bearing directly on the distribution of intelligence, we should, I think, assign the greatest antecedent probability to the *second* of the foregoing hypotheses, since, as Pearson himself (1895) has pointed out, the prevailing curve in zoological and anthropological material is type IV with 'a moderate degree of skew variation'. This is the curve I myself originally assumed (1917, p. 34; cf. 1937, p. 442 and refs.).

III. The empirical evidence

Let us now turn to the objective evidence furnished by first-hand surveys and the like, and inquire how far it tends either to confirm or to confute these rival conjectures. As Mr Lewis observes, the practical experience of school teachers and education officials strongly suggests 'a humped or bell-shaped distribution', but throws no light on the problem of symmetry. 'Most important of all', therefore, in his view, 'is the claim that the normal distribution of intelligence has been established from the results of intelligence tests: e.g., Burt, 1953'. As worded, this statement may perhaps be a little misleading.

To begin with, the data on which I myself have principally relied consist, not so much of 'raw scores' or 'quotients derived from the Binet scale' and taken at their face value, but rather of corrected assessments based on all the information available. What we are hoping to determine is the distribution of *innate* ability; and the results of a single test of intelligence applied on a single occasion are too much affected by environmental conditions to be accepted for this purpose just as they stand. It has therefore been my regular practice to submit to the teachers in the schools concerned a list of marks in order of merit for pupils in their charge, and to ask them to criticize the actual scores; then, whenever any marked discrepancy has emerged, the child has been re-examined, usually with a different type of test (1943, p. 98).

Secondly, my 'claim' was not to have 'established the normal distribution of intelligence', but rather the contrary. From 1917 onwards the normality of the frequency-distributions obtained from my L.C.C. surveys was regularly checked by the χ^2 method (cf. Burt, 1917, p. 34); and in every case the distribution was found to diverge significantly from the normal. Certainly, for many practical purposes the distribution of pupils in the ordinary schools (i.e. excluding specific cases of deficiency) might, as I pointed out, be treated as '*approximately* normal, but only *approximately*'; and to avoid misunderstanding I italicized the word (1943, p. 92).

However, if I understand them aright, Mr Lewis, Mr Richmond, and Dr Heim would deny even a tendency towards approximate normality. Any such approximation, they hold, is a sheer artefact—a by-product of the arbitrary way in which the tests have been standardized. Similarly, Mr Richmond, as we have already seen, declares that the psychometrist can only secure this appearance of normality by tinkering with the material that his tests supply; and 'this rigging of the test-situation', he insists, 'is as false as it is artificial' (1953, pp. 227f.).

Mr Richmond makes these assertions dogmatically without explicit evidence. But Mr Lewis goes further: he attempts to demonstrate them. His argument is a little difficult to follow, because, instead of offering a general algebraic proof, he deduces his conclusions from five artificially constructed examples. For each he chooses a test with six items and a sample of six individuals.* But with such small numbers it is impossible to prove any general proposition about frequency distributions. *With the particular numbers he selects*, Mr Lewis is able to show that, when the items are all of equal difficulty ($F.v. = \frac{1}{2}$), 'central concentration' (the nearest he can get to approximate normality) can only occur when the validity coefficients (E_{13}) range from 0 to 1; if, however, all or nearly all the items have perfect validity ($E_{13} = 1$), then there will be 'a tendency towards rectangularity rather than normality'. Now, it is quite easy to demonstrate that, with a different number of items or a different number of testees, precisely the opposite may occur. Suppose the test contains seven items instead of six, then, with the conditions Mr Lewis specifies for a 'tendency towards normality' (scheme 1, p. 100), we obtain the following:

Successful groups	*a* *b* *c*	*a* *b* *d*	*a* *b* *d*	*a* *b* *e*	*a* *c* *d*	*a* *c* *e*	*b* *c* *f*
$F.v.$ E_{13}	$\frac{1}{2}$ 1	$\frac{1}{2}$ 1	$\frac{1}{2}$ 1	$\frac{1}{2}$ $\frac{1}{2}$	$\frac{1}{2}$ $\frac{1}{2}$	$\frac{1}{2}$ 0	$\frac{1}{2}$ 0

Frequency distribution

f	*e*	*d*	*c*	*b*	*a*
1	2	3	4	5	6

Here, instead of a 'central concentration', we now get a specimen of 'rectangular' distribution quite as perfect as any Mr Lewis can produce with the conditions which he himself prescribes (scheme iv, p. 101).

But I need not labour these objections, since no statistical psychologist, I fancy, would wish to dispute Mr Lewis's final conclusion: namely, that the frequency-distribution obtained from a composite test can be forced into almost any shape or form by suitably selecting (or scaling) the constituent test-items, provided the selection (or scaling) is based *on the same sample of persons*. It is merely his proof, and his formulation of the methods needed to secure this or that type of curve, that are incorrect (for a more rigorous approach to the same problem, see Lord, 1952).

It seems clear, however, that such methods will no longer serve when the individuals or groups used for the final distribution are drawn from a type of population *different* from that which is used for selecting or scaling the items. This will become obvious if we consider Mr Lewis's fourth scheme (p. 101). Here we might imagine his six groups, *a*, *b*,…, *f*, to be six consecutive age-groups, and his six items, 1, 2,…, 6, to be six ideal questions of the Binet-Simon type, each a little harder than the last and each passed by the same percentage of children of the age for which it is designed (Mr Lewis assumes 100%, which would be impossible in practice; I should prefer 50%). Then, as his scheme implies, the average mark obtained by the individuals in each *successive* age-group would yield a rectangular distribution. That of course is true of the actual Binet tests. But his argument in no way demonstrates that the intelligence of the individuals within the *same* age-group is distributed in rectangular fashion; it might just as well conform to some bell-shaped curve, possibly normal, possibly not: we cannot possibly tell until we have examined the actual scores.

---

* Mr Lewis calls them 'equal-sized groups', and assumes that each member in any one group responds identically throughout the series—an assumption which could never be fulfilled unless the testees were as rigorously selected as the items.

In point of fact, as Mr Lewis implies, he is really thinking of the methods adopted for selecting and evaluating test-items in a *group* test such as the Moray House. Now the purpose of such tests is entirely practical: they were never intended to determine the theoretical distribution of intelligence, but merely to enable the users to allocate pupils of one particular age to the most suitable school or class. Moreover, unlike the Binet tests, they are standardized solely on children of the age for which they are to be used. Quite illogically, however, having shown that the distribution of measurements from Moray House tests is 'imposed on the scores during the process of standardization', Mr Lewis goes on to infer that the distributions obtained with the Binet tests will similarly depend on the methods of selecting and scaling the constituent items: 'its bell-shaped distribution of quotients, no less than the roughly rectangular distribution of Moray House Tests, may be regarded as a consequence of the "facility" and of the E_{13} values of the component items'. But, since the Binet scale is standardized in an entirely different fashion, this is a complete *non sequitur*.

(i) *Scaling by mental age.* Mr Lewis deals with the method of selecting items for the Binet scale in another section of his paper. Binet, we are told, allocated a test-item to a given age 'if it could be passed by about two-thirds of the testees *of that age*: Burt (1921), in his adaptation of the Scale, adopted instead the ratio one-half'. This implies that, for Binet's ratio of two-thirds of the specific age-group, I myself merely substituted the ratio of one-half of the *same age-group*. But that is incorrect. My proposal was based on an argument quite different from Binet's. The guiding idea was that the age level assigned to any item could be regarded as a 'threshold' or 'limen' (in the sense of the theory of psychophysical methods). The working procedure was laid down as follows: 'Let the children be grouped, *not*, as with Binet, by the age at their nearest birthday, but by their age at their *last* birthday. The age of the groups will thus average 5½ years, 6½ years, . . ., respectively, not 5·0 years, 6·0 years, ... Now a correct allotment of the test implies that in every group the average mental age shall coincide with the average chronological age. Hence to be assigned to any given year (as a teacher would interpret a mental year*), a test should be passed by approximately 50% of the children of that age.' In short, the measurement 'x mental years' means the performance of a *median* child of x chronological years (Burt, 1921, pp. 140 and 271).†

---

* If, for example, *all* the children aged 5·0 to 6·0 pass all the tests for age 4·0 to 5·0 and *half* the tests for age 5·0 to 6·0, both their chronological and their mental age will average 5½: from this it is clear that the same requirement will be fulfilled if we suppose that *half* the children pass *all* the tests for age 5·0 to 6·0

† The accuracy of this principle seems to have been endorsed by most later psychometrists. Guilford (1936, pp. 210–1) discusses the appropriate procedure for Binet tests in his chapter on 'Constant Methods'. 'In establishing the age levels of his various tests, Binet' (he writes) 'adopted a rough and ready standard: any test that could be passed by from two-thirds to three-fourths of the children of a certain age should be placed at that age level. This means that the average child can do *more* than pass the tests for his age. This is a violation of the use of medians and means. A five-year test should be one that the average five-year-old child can pass, or one that 50% of them can pass.' For the Terman–Merrill revision, a similar principle was subsequently adopted (Terman and Merrill, 1937, p. 9). For the original Stanford Revision a rather different procedure was used.

Now how can the use of medians for *successive* ages affect the distribution of mental ages for individual children *within* a single year? No assumption of normality is made; and the fact that the distribution actually secured fails to conform with the normal clearly proves that the mode of selection does not necessarily entail normality. All that is assumed in the process of scaling is that the intervals of difficulty from one mental age to another may be treated as equal.

To illustrate the type of distribution obtained, I give the results of one of the most thorough of the London surveys (Fig. 1). The figures are based on the results of the English Revision of the Stanford–Binet Scale, checked by the teachers; the children tested, 2835 in number, formed an unselected sample of the general school population between the ages of 6 and 11. The data were collected in the course of a joint research on the standardization of the Stanford–Binet Scale for English children (Burt, 1953, p. 348). I must again express my gratitude both to my co-workers and to the head-teachers concerned for their willing collaboration. The survey was organized as part of my regular duties for the L.C.C.

In terms of the I.Q., the mean is 100·3, the median 101·0, and the standard deviation 14·9 (15·2 for the lower half, 14·7 for the upper). The dotted line shows the best-fitting normal curve. On applying Pearson's test for goodness of fit, we obtain $\chi^2 = 59{\cdot}1$, and $P = 0{\cdot}000001$; i.e. it is about a million to one against the sample being drawn from a normally distributed population. The functions required for determining the form of the curve (cf. Pearson, 1914, p. lxiii) are $\beta_1 = 0{\cdot}035 \pm 0{\cdot}022$, $\beta_2 = 3{\cdot}36 \pm 0{\cdot}18$, $\kappa = 0{\cdot}04$; this implies that the curve is of type IV. The skewness $(-0{\cdot}079 \pm 0{\cdot}024)$ is small but statistically significant; the leptokurtosis less marked but still fully significant.* If from the distribution here depicted specific cases of mental deficiency are omitted, χ^2 drops to 26·7 and P rises to 0·044. Thus the deviation from normality still remains statistically significant, though it would only be demonstrable with fairly large samples.

Dr. Heim and Mr Lewis, as we have seen, attribute all appearances of normality to the skill of the test constructor; but other critics may be inclined to wonder whether *deviations* from normality, such as we discover here, are not merely the incidental consequences of his *lack* of skill. This interpretation might seem plausible had we had no more than one or two sets of data to account for. But the numerous surveys carried out by various investigators nearly all yield constants of much the same order, generally tending in the same direction. As I have pointed out elsewhere, 'for most of the distributions β_1 lies between 0·0 and 0·2, and β_2 between 2 and 4', so that the skewness usually lies between 0·05 and −0·10 (Burt, 1943, p. 93). This general agreement indicates some more fundamental cause, such as is suggested by the second of the four hypotheses mentioned above.

---

* It is instructive to compare this with the corresponding values for bodily measurements. For male height the British Association figures (1883) give $\beta_1 = 0{\cdot}0002$, $\beta_2 = 3{\cdot}15$, $sk = -0{\cdot}01$, but probably many dwarfs were omitted from the survey; for weight, $\beta_1 = 0{\cdot}54$, $\beta_2 = 4{\cdot}46$, $sk = +0{\cdot}29$ (Yule & Kendall, 1937, pp. 161, 546). Very similar beta-functions are given for skeletal age (eleven age-groups) by Dearborn & Rothney (1941, p. 103). In testing normality I myself should prefer to use estimates of the gamma-functions, as suggested by Fisher (1934, pp. 53, 77); but for bodily measurements the beta-functions have hitherto almost invariably been adopted. As I have pointed out elsewhere, with large and unselected samples the most suitable equations for fitting a type IV curve to data for intelligence seem to be those for the hypergeometric series. The normal curve provides a limiting case; its equations are therefore simpler, and will yield reasonable approximations, except where frequencies for markedly subnormal or supernormal children are required: for such satisfactory estimates may be obtained most simply by assuming that the curve consists of two normal halves with slightly different standard deviations. These devices have been so often used in practice (cf. Burt, 1943, p. 85 and refs.) that it is surprising Mr Lewis makes no reference to them.

With tests of attainments, the distributions obtained are of a rather different type (cf. Lord, 1954).

Even where the deviations from normality are not fully significant, they are still of the same nature as those described above. One of the largest surveys of unselected children is that carried out by Dr Fraser Roberts and his colleagues (1935, pp. 334f., 'Form of Distribution of Scores'). Practically every child in the city of Bath between the ages of 9 and 13—2553 in all—was tested with the Advanced Otis Scale. With the χ^2 test, the odds against the data being drawn from a normally distributed sample were about 6 to 1. The divergence therefore is not fully significant. But there is a definite 'heaping up at the lower end of the scale'; and this, as the authors observe, is 'typical of the results of intelligence testing' (cf. also Fraser Roberts, 1940, pp. 237f.).

However, the data supplied by the Binet tests is only part of the evidence available for examination. Tests scaled in entirely different ways point, as we shall see, to much the same conclusion.

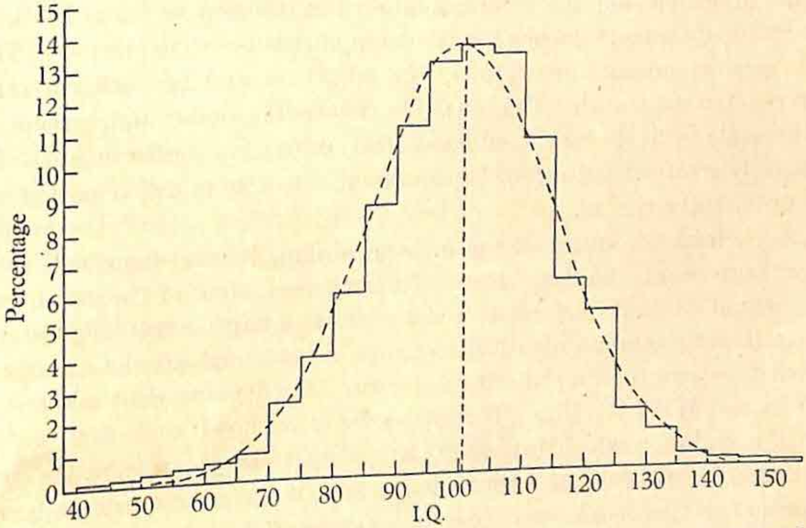


Fig. 1. Distribution of intelligence. The dotted curve shows the best fitting normal distribution.

(ii) *Scaling by just noticeable intervals or equal appearing intervals.* In the early days of intelligence testing, it seemed doubtful how far we were justified in assuming that the difference between one mental age and the next was the same throughout the range concerned. I myself (1914, 1921) believed that with the Binet Scale the intervals showed only a very rough equality. Accordingly, alternative methods of scaling were tried—most of them based on the traditional psychophysical procedures. For instance, in constructing a test of Opposites, Mr Moore and I arranged a large number of items in order of difficulty, and then submitted successive pairs to twenty teachers, with the following instruction: 'Will you compare the difficulty of these two items, and say whether you consider the second to be (i) harder, (ii) easier, or (iii) of about equal difficulty?' Whenever approximately 50% of the judges agreed that the second was harder than the first, we regarded the difference as 'a just perceptible increase'; and on this basis we picked out a series in

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated ....
Accs. No ....

which the increment of difficulty exhibited by each item over its predecessor was 'just perceptible' in the sense defined. Where fewer items were available or needed (e.g. with Completion tests), we graded them by employing instead the method of 'equal appearing intervals'. On applying tests so standardized to random samples of children, the distributions obtained were always approximately normal, though never exactly so. Similar results were reported in other countries (cf. the curves printed by Trabue, 1916, figs. 10–14: for the psychophysical methods, v. Brown & Thomson, 1925, pp. 46f).

(iii) *Scaling by number of unit-processes performed.* All the foregoing are 'graded' tests. In what may be called 'homogeneous' tests the constituent tasks are all of approximately equal difficulty. With these the examinee's score is expressed in terms of the number of unit processes correctly performed within a given time limit.

There are numerous varieties. Before the current type of group tests attained their present vogue, one of the most popular devices was the so-called erasure or cancellation test. The examinee is given a sheet of pied print, and is asked to cross out one or more letters (A, E, O, for instance) wherever they occur: the measure is simply the number of letters correctly cancelled. These and other tests, scored on a similar basis, were freely used by earlier investigators; and frequency distributions may be found in their publications or unpublished theses:* graphs for the commoner types were printed in Thorndike's first book on mental measurement (1904, figs. 30, 31 32, and 34; each diagram shows a normal curve fitted to the distribution). As Thorndike points out, except when the nature of the scale is likely to introduce distortion (e.g. in measurements of reaction-time), practically every distribution 'approximates to the so-called normal shape, i.e. that of the probability curve'.

Thus, with each of the different methods of scaling I have described, the resulting distributions have much the same form. Nevertheless, none of these scaling methods would be accepted as final by the geneticist without a further test. For him a scale of measurement is satisfactory only if the effects of each and all the various causative agencies—whether genes or non-inheritable factors—are demonstrably additive (cf. Fisher *et al.* 1932; Sewall Wright, 1952). This of course is an ideal; and often we have to be satisfied with a scale for which the effects are additive *on the average.* In dealing with animals this can be checked in various ways: e.g. if we cross two true-breeding lines (P_1 and P_2 say) and then backcross the first generation (F_1) to one or other of the parent groups, we shall obtain means for five groups, $\overline{P}_1$, $\overline{B}_1$, $\overline{F}_1$, $\overline{B}_2$, $\overline{F}_2$; and, if the scaling is additive, these should be equidistant; if the contributions to the variances are also additive, the variances of the true-breeding lines and of the first generation should be the same, even when the means differ (for an admirable application of this procedure see Chai, 1956). With physical measurements, so it appears, the natural unit fulfils these requirements well enough in the majority of cases; but occasionally (e.g. with body weight or volume) the original measurements need to be re-scaled by adopting a logarithmic or similar transformation. With psychological measurements and with human beings it seems possible that rough tests along analogous lines might be devised by selecting appropriate pedigrees. For intelligence, as we have just seen, much the same type of distribution is reached with several independent methods of scaling. Hence it seems fair to infer that here too what may be regarded as 'natural units' are adequate for the broader issues with which we have been concerned.*

---

* A description of tests of this type which had the highest validity will be found in a joint paper by Mr R. C. Moore and myself (1912): frequency distributions were given in Mr Moore's thesis.

However, the study of frequency distributions forms by no means the only source of evidence available for verifying the hypothesis proposed: the coefficients of correlation, the bivariate distributions for parents and their offspring, and for children and their brothers and sisters, all tend to confirm the same theory; and the other statistical constants all closely agree with the theoretical values deducible from the hypotheses outlined above (Burt, 1935). In the revised version of his article, Mr Lewis briefly refers to the paper in which these further arguments are set forth; but he makes no attempt to rebut them.

Finally, there is the more practical test. The assumption of approximate normality has now been adopted for a wide variety of administrative purposes over a long period of years. In organizing classes, in planning the number of schools to be built or allocated, in devising or standardizing schemes of marking or grading, and in numerous other ways, the scheme of proportional frequencies suggested by the normal curve (e.g. those of the L.C.C. *Report*, Burt, 1917) has proved both serviceable and reliable, despite all the objections urged by the earlier sceptics forty years ago. It is difficult to understand how this could have happened if the foundations on which such schemes are based were as faulty as the modern critic avers.

IV. SUMMARY AND CONCLUSIONS

In current controversies about the distribution of intelligence, the term intelligence is interpreted in very different ways. In this paper it is used in the technical sense given to it, explicitly or implicitly, in the work of Spencer, Galton, Binet, and their followers, namely, 'the innate general factor underlying all cognitive activities'.

Four main hypotheses have been advanced to account for the manner in which it is distributed; and the frequency-curves deduced from each of the hypotheses prove to be sufficiently distinctive to furnish a crucial test. Of the four rival theories that which, on genetic grounds, has the strongest *a priori* arguments in its favour assumes that intelligence, like most other graded characteristics, is the result of both multifactor and unifactor inheritance. This implies as a corollary that the curve of distribution will be bell-shaped and moderately asymmetrical: i.e., 'type IV' rather than the 'normal' type.

---

* Space will not permit me to reply to Mr Lewis's criticisms of the I.Q. as a unit of scale. He quotes Dr Banks and myself as 'claiming that I.Q.'s should be interpreted as measurements on an interval scale, so that the differences between I.Q.'s may be legitimately compared', and making the 'still more ambitious claim that with Thurstone's technique the measurements may be interpreted as measurements on a "ratio scale"'. 'The basis of these interpretations' (he adds) 'is the normal distribution'. Now, if the reader will turn to the page Mr Lewis cites (Burt, 1953, p. 161), he will see that I.Q.'s and Thurstone's scale were inserted in our table merely to *illustrate* the concepts of an 'interval scale' and a 'ratio scale' in psychology: not a word was said about 'claims', or about the 'legitimacy' of such procedures, or about the normal distribution as a 'basis'. On the contrary, the point of the table was to show that measurement in psychology need *not* necessarily take either of these two forms. Mr Lewis's argument in the paragraph which precedes these criticisms—namely, that those critics who suppose that all psychological measurement must be based either on an 'interval scale' or a 'ratio scale' have overlooked the use and fundamental nature of the 'ordinal scale'—is precisely the point that we ourselves were concerned to make (*loc. cit.*, last paragraph).

Dr Heim similarly criticizes psychometrists for their 'universal' and 'unhesitating' use of the I.Q., and recommends a percentile basis instead. But percentiles (or S.D.'s derived from percentiles) were in fact what I myself had recommended and used for more technical inquiries (cf. *The Backward Child*, pp. 20–24 and refs.); and I have repeatedly stated that I do *not* consider the I.Q. to be the best unit for scientific purposes (e.g. 1917, p. 15). However, these more elaborate procedures have been assailed by critics on the opposite flank—those who dislike all 'tinkering' with the initial data. For the graph printed here, therefore, I have retained the I.Q. as being at once the most direct and the most intelligible unit. But I should emphasize that I do not consider the units, the mean, or the standard deviation, used to compute the smoothed curve in this particular survey, to represent the most appropriate measurements for other problems besides that examined here (cf. Burt, 1937, p. 442). And, should any reader feel that they may be equally inappropriate for the present problem, may I point out that the majority of the graphs and frequency-curves cited above were *not* based on the I.Q.?

The empirical evidence is derived principally, though not entirely, from the results of intelligence tests. Except for those intended solely for a single age-group, most tests are scaled by methods independent of any presupposition about the resulting distribution: e.g. (i) by mental age, (ii) by least perceptible or equal-appearing intervals, or (iii) by the number of unit processes performed. With each of these procedures the frequency-curves obtained yield, in nearly every case, statistical constants implying a definite though slight degree of leptokurtosis and negative asymmetry. Thus the results are fully consistent with the twofold hypothesis suggested by genetic arguments, and entirely incompatible with the other three hypotheses.

If children certified as mentally deficient under the Acts of 1913 and 1921 are omitted—a procedure which eliminates most of the cases due to unifactor inheritance—then the divergence from normality is comparatively slight, and with small samples ceases, as a rule, to be statistically significant. However, in computing the frequencies of well-marked deviations, whether in the direction of high ability or low, any assumption of strict normality should be used with the utmost caution.

REFERENCES

BANKS, C. & BURT, C. (1953). Statistical analysis in educational psychology. *In* Mace, C. A. & Vernon, P. E. *Current Trends in British Psychology*. London: Methuen.

BRITISH ASSOCIATION (1883). Report of anthropometric committee. *Annual Report*, pp. 256–73.

BROWN, W. & THOMSON, G. (1925). *The Essentials of Mental Measurement*. Cambridge University Press.

BURT, C. (1911). Experimental tests of higher mental processes and their relation to general intelligence. *J. Exp. Pedag.* **1**, 92–112.

BURT, C. (1912). The inheritance of mental characteristics. *Eugen. Rev.* **4**, 1–33.

BURT, C. (1914). The measurement of intelligence by the Binet tests. *Eugen. Rev.* **6**, 36–42, 140–53.

BURT, C. (1917). *The Distribution and Relations of Educational Abilities*. (L.C.C. Reports, no. 1868). London: P. S. King.

BURT, C. (1921). *Mental and Scholastic Tests*. (L.C.C. Reports. no. 1925–7). London: P. S. King.

BURT, C. (1935). *The Subnormal Mind*. Oxford University Press.

BURT, C. (1937). *The Backward Child*. London: University of London Press.

BURT, C. (1943). Ability and income. *Brit. J. Educ. Psychol.* **13**, 83–98.

BURT, C. (1955*a*). Galton Lecture. *Eugen. Rev.* **47**, 81–91.

BURT, C. (1955*b*). The evidence for the concept of intelligence. *Brit. J. Educ. Psychol.* **25**, 158–77.

CHAI, C. K. (1956). Analysis of quantitative inheritance in mice. *Genetics*, **41**, 157–78.

DARLINGTON, C. D. & MATHER, K. (1949). *The Elements of Genetics*. London: Allen and Unwin.

DAVIES, H. (1941). *The Analysis of Economic Time Series*. New York: McGraw-Hill.

DEARBORN, W. F. & ROTHNEY, J. W. (1941). *Predicting the Child's Development*. Cambridge, Mass.: Sci-Art Publishers.

FISHER, R. A. (1918). The correlation between relations on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinb.* **52**, 399–434.

FISHER, R. A. (1934). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd.

FISHER, R. A. *et al.* (1932). The genetical interpretation of statistics in the study of quantitative inheritance. *Genetics*, **17**, 107–24.

FRASER, D. C. (1954). *A Psychological Glossary*. Cambridge: Heffer.

GUILFORD, J. P. (1936). *Psychometric Methods*. New York: McGraw-Hill.

HEBB, D. O. (1949). *The Organization of Behaviour*. New York: Wiley.

HEIM, A. (1954). *The Appraisal of Intelligence*. London: Methuen.

KEYNES, J. M. (1921). *Treatise on Probability*. London: Macmillan.

LEWIS, D. G. (1957). The normal distribution of intelligence: a critique. *Brit. J. Psychol.* **48**, 98–104.

LORD, F. M. (1952). The relation of test score to the ability underlying the test. *Psychometr. Monogr.* **7**.

LORD, F. M. (1954). A survey of actual test-score distributions with respect to skewness and kurtosis. *Educational Testing Service Research Bulletin*. Princeton University Press.

MOORE, R. C. & BURT, C. (1912). The mental differences between the sexes. *J. Exp. Pedag.* **1**, 273–91, 355–72.

PEARSON, K. (1895). Skew variation in homogeneous material. *Phil. Trans.* A. **186**, 343–61; **187**, 443–59.

PEARSON, K. (1903). On a generalized theory of alternative inheritance with special reference to Mendel's laws. *Phil. Trans.* **203**, 53–87

PEARSON, K. (1914). *Tables for Biometricians and Statisticians*. Cambridge University Press.

RICHMOND, W. K. (1953). Educational measurement: its scope and limitations: a critique. *Brit. J. Psychol.* **44**, 221–31.

ROBERTS, J. A. F., NORMAN, R. M. & GRIFFITHS, R. (1935). Studies on a child population. *Ann. Eugen., Lond.*, **6**, 319–38.

ROBERTS, J. A. F. (1940). *An Introduction to Medical Genetics*. Oxford University Press.

SNYDER, C. (1940). *Capitalism the Creator*. New York: Wiley.

TERMAN, L. M. & MERRILL, M. A. (1937). *Measuring Intelligence*. London: Harrap.

THORNDIKE, E. L. (1904). *Introduction to the Theory of Mental and Social Measurements*. New York: the Science Press.

TRABUE, M. R. (1916). *Completion Test Scales*. New York: Columbia University Press.

WRIGHT, SEWALL (1952). Genetics of quantitative variability. *Quantitative Inheritance*. London: H.M. Stationery Office.

YULE, G. U. & KENDALL, M. G. (1937). *Introduction to the Theory of Statistics*. London: Griffin.

(*Manuscript received 4 July 1956*)

Note on the 'Index of Item Validity'. Presumably because he believes that intelligence follows a rectangular distribution, Mr Lewis (loc. cit., p. 99) has preferred to assess item validity by a point biserial correlation instead of the more usual tetrachoric correlation. In that case with three equal subgroups the equation should be (in his notation)

$$E_{13}=\frac{3(S_1-S_3)}{n}$$

This is the formula I have used for the table on p. 168. Mr Lewis's equation as printed substitutes a multiplication sign for the subtraction sign. But probably this is a slip either in the algebra or in the printing. (See corrigendum below—Ed.)

CORRIGENDUM

In 'The normal distribution of intelligence: a critique' by D. G. Lewis, Vol. **48**, Part 2, the formula on p. 99 should read as follows:

$$E_{13}=\frac{S_1-S_3}{\frac{1}{3}n}.$$

ON THE VALIDITY OF APPLYING COMMUNICATION THEORY TO EXPERIMENTAL PSYCHOLOGY*

By E. COLIN CHERRY

Imperial College, University of London

Communication theory is properly applied to human performance only under certain conditions: for example, the application must be made by an outside person and not by the subject himself. The theory also requires averages of numerous events rather than dealing with isolated ones. Given these and similar conditions, some useful and valid applications may be made to psychology. It is perfectly possible to apply the theory in the case in which stimulus and response are continuous functions of time, although this is not always realized.

I. Introduction

We shall discuss briefly some of the difficulties of using the formal Mathematical Theory of Communication (in the sense of Shannon, 1949) as an aid to constructing and interpreting experiments involving human behaviour. Note that the title refers to '...validity of *applying*...'; it is not the validity of the mathematical theory itself which is here in question, but only the suitability and correctness of its application to certain types of experiment. My remarks will be illustrated with one or two well-known psychological experiments, as models. It will be assumed that the reader is already familiar with the principal methods and theorems of Communication Theory.

We can observe a human operator or two humans forming a communication link (as, for example, a conversation) only by setting up another channel—the *meta-channel*, or *channel of observation* (Fig. 1).

The Shannon Theory of Communication is written in the language of this observer. It is built upon the concepts of a message source ensemble, with message probabilities $p(x)$ and a noisy† received signal ensemble, with probabilities $p(y)$. It uses also conditional probabilities, such as $p_y(x)$ and others. Now these probabilities can be assessed only by an observer who can observe both the transmitted signs x and the received noisy signals y. That is, by an observer *who can see both ends of the channel*—an external observer, as in Fig. 1. Expressions such as x, y, $p(x)$, $p(y)$, $p_y(x)$, etc., are then proper only to the meta-language of such an external observer and not to that of the receiver who knows only the received signals y. The recipient of the signals can observe only those signals y and estimate their probabilities of arrival; he cannot observe the sent message x except inasmuch as he can regard y as evidence about them. But he cannot directly observe x, nor can he estimate probabilities $p(x)$.

In physical, including psychological, experiments, observations are made and reported upon; and the nature of the reports, the theory set up, will depend upon whether the observer is completely outside the phenomenon he observes, or whether he himself is part of it. For example, in astronomy the observer, by looking through his telescope, does not affect the stars in their courses. By contrast, in many laboratory experiments the observer may disturb the observed phenomenon by his very observations; or he may need to make changes, or alter some situation, before the phenomenon comes to light. But experiments upon 'communication' are particularly prone to disturbance by the observer as when for instance, this observer is one of the communicants. By an external observer then we mean one who merely watches as through a peephole and records; he makes no disturbances once his experimental situation is set up. We shall later discuss the consequences of confusing these two kinds of observer. In psychological experiments we are dealing with a unique class of phenomena—human beings, who are themselves articulate and can make their own reports. Both languages have their uses; that of the external observer and that of the human subject under observation. But, as stressed already, Shannon's Communication Theory is written wholly in the meta-language of an external observer; he himself exhibits this fact by figure 8 of his book (1949), and by the theorems which he derives from the assumptions shown in that figure.

---

* A revised version of a paper read at a Symposium on Human Operator Problems, organized by the Ministry of Supply, 19 September 1955, London.

† By 'noisy signals', we mean signals which are subject to disturbances which are not known in moment-by-moment detail, but only statistically, e.g., random disturbances.

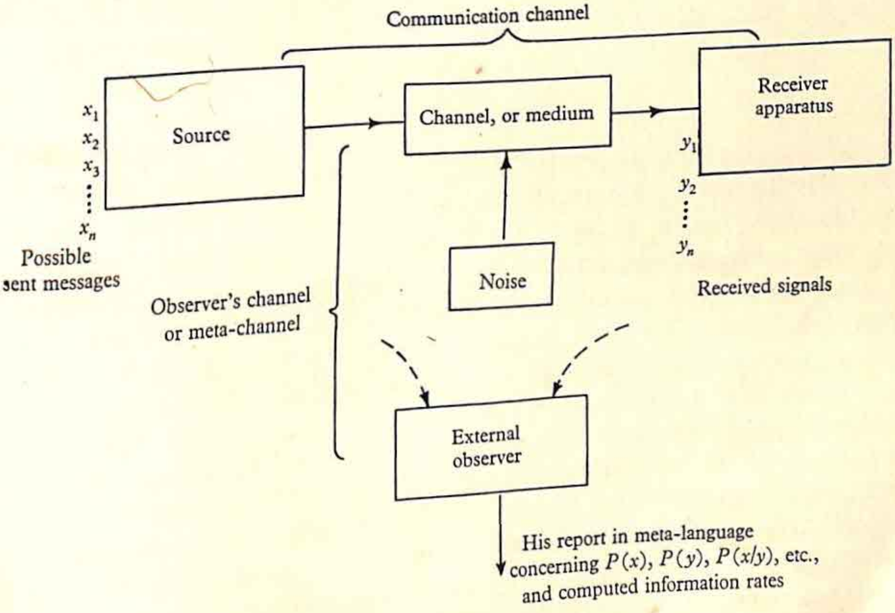


Fig. 1. A communication channel and its external observer.

II. THE BASIC ASSUMPTIONS OF SHANNON'S COMMUNICATION THEORY

I shall start with a bare statement of the basic assumptions of Shannon's mathematical theory and then proceed to examine them, in relation to the design of human operator experiments. The theory was first set up to describe the properties of a communication source, channel and receiver, as exampled, say, by a teleprinter channel; such a system is not precisely analogous to a human operator with his sense-stimuli and overt responses. It will therefore be helpful to distinguish between (*a*) a communication channel and (*b*) a human operator. Let us list the assumptions of the theory as relating, most properly, to the former.

(A) The theory is expressed entirely in the (meta-) language of an external observer, not in that of one of the participants (Fig. 1). We observe a communication channel only by setting up another—the 'Observation Channel'.

(B) Shannon's theory is a statistical one; that is, it concerns only average, macroscopic, properties of a system, not specific momentary properties.

(C) From long series of past observations it is assumed that the external observer has made good estimates of the probabilities (frequencies) of occurrence of the various transmitted message signals x, received signals y and interfering noise. Further, it is assumed that the noise source(s) can be identified ('singled out') and their statistical properties specified by the observer; or, in certain simple cases, that noise is absent.

(D) It is assumed that, *a priori*, the two communicants operate with an agreed set of message signs, or alphabet, having certain definite frequencies of occurrence. It is assumed that these frequencies remain unchanged with the passage of time; that is, that the statistics of the situation are stationary.

(E) It is assumed that this alphabet of signs $x_1, x_2, \ldots, x_n$ (signals, ensemble of messages, coding system) remains unchanged. (We beg the question of so-called 'continuous' information sources for the moment.)

III. THE HUMAN OPERATOR ('SUBJECT') COMPARED TO A COMMUNICATION CHANNEL

Comparison between a human operator and a communication channel is to-day commonplace; the stimuli represent 'input signals' applied to his sense-organs, while his overt responses represent 'output signals'; the human operator represents the 'transducer'. But there are a few essential points to be borne in mind, when using this analogy:

(*a*) A human operator is an integrated organism, and his ears, eyes, nose, finger-tips, etc. are not like independent input terminals; the responses made depend upon the conditions at all the sense organs, which are mutually dependent. What a man thinks he can smell depends upon what he sees, and so on.

(*b*) A human being is an ensemble of experiences; as a 'transducer', his parameters, or properties, may change after every communicative experience (i.e. perception). Again, he cannot be represented by one set of fixed parameters, because these may depend upon the nature of the stimulus itself and of the environment.

(c) Every individual represents a different ensemble of experiences, so that input stimuli become associated with very varied sets of prior stimuli in determining responses; that is, the 'meaningfulness' of stimuli is a fluid and varied property of human information-transfer, rendering it difficult to make the choice of stimuli sufficiently universal, in behavioural experiments. We may try to smooth out the variations by choosing groups of people of similar educational background, in a close age-group, etc. Again for stimuli we try to choose letters, words, signs of various kinds, such as may be assumed to be 'widely known'. But Communication Theory, in Shannon's terms, makes no reference to 'meaningfulness' but only to *syntactic* properties and rules—only to signs and their probabilities, joint probabilities, etc., as assessed by the external observer. Therefore:

(*d*) If we choose to ignore the 'meaningfulness' of stimuli, but rather to regard stimuli, like letters, words, etc., purely syntactically as some philosophic schools will have (to which, incidentally, I do not belong) then we really are assuming that the responses of a human operator to these stimuli will depend only upon the range of choice they offer and their probabilities of occurrence. In the experiments of Hick (1952, 1953), for example, this assumption is reasonably justified, since his stimuli are provided visually by a simple ring of lamps, the 'meaningfulness' of which will surely be fairly universal and of little influence. In his examples, Communication Theory gives a good, valid interpretation. But, furthermore, Miller (1951) cites several examples of stimulus-response experiments using letters and words, but in which the results are interpretable, objectively and repeatedly, in terms only of probabilities, without reference to 'meaningfulness'.

It seems to me to be one of the major discoveries of recent behavioural experiments, that responses to linguistic stimuli are interpretable this way; certainly some credit for this is attributable to the discipline of statistical Communication Theory. But more on this later.

(*e*) We have mentioned 'probabilities of stimuli'. Probabilities to whom? If we take any one specific person (human operator) we cannot speak of the 'probability' of a certain letter or word to him, as a frequency; doubtless, such probabilities are represented, matrix-wise, in the structure and organization of his brain—but they are not assessable, either externally, or by the operator himself. Probabilities, to him are 'degrees of belief' and at best may be rank-ordered, but are non-numerical. But the Theory of Communication takes probabilities as relative frequencies; hence:

(*f*) If we use linguistic stimuli, we may use the relative frequencies of the letters and words 'in the language' as the probabilities, and then average our results over large numbers of human operators drawn from this language group. In contrast to Communication Theory, which usually refers to an ensemble of messages, we often require also an ensemble of operators (subjects).

(*g*) The difficulty of 'non-stationariness', corresponding to learning ability, may be dealt with in the same way. In a communication channel, we may assume the apparatus has fixed, definable, properties like bandwidth, noise-level, modulation limits, etc., but does not change its structure with time and 'experience'. Communication Theory then commonly treats *time*-average statistics of input and output signals. But human operators learn, and conditions of test may not necessarily remain stationary with time; however, we may nevertheless take *ensemble*-averages, on the assumption that an ensemble of operators will change, or learn, in nearly-like manner, under the same test conditions.

(*h*) One most important aspect of the human operator in comparison to a communication channel, which arises when Communication Theory methods are used, is easily overlooked. The human operator, the organism, cannot be 'taken apart'. Even if regarded theoretically, as a 'black box' transducer having an unspecified arrangement internally, it it questionable whether we can validly subdivide its apparent functions and discuss them separately, as we commonly do in the case of the various parts of an electrical communication channel.

To illustrate these important points (*a*)–(*h*) it is convenient to return to our distinction between the external observer and the participant observer, referring to Figs. 1–3.

External observer and participant observer

The external observer (Fig. 1), as assumed by Shannon's Communication Theory, can observe both ends of the communication channel. He observes the transmitter's alphabet of message signals $x_1, x_2, \ldots, x_n$ and assesses their relative frequencies $p(x_1), \ldots, p(x_n)$; he observes the noise 'signals' and the received signals $y_1, y_2, \ldots, y_n$ similarly. If we imagine the channel as being, say, a long length of telegraph cable, and the 'receiver apparatus' as a teleprinter, it is sometimes convenient to treat these separately. For example, the received signals y may be observed as noisy signals received at the end of the cable, say at the terminals of the detector in the receiving apparatus; the observer may then compute the transmitted information rate $H(x)$ and the equivocation $H(x/y)$ (average ambiguity about x existing potentially in received signals y), thereby giving the rate of transmission of information R between the source and the receiver detector terminals. Here:

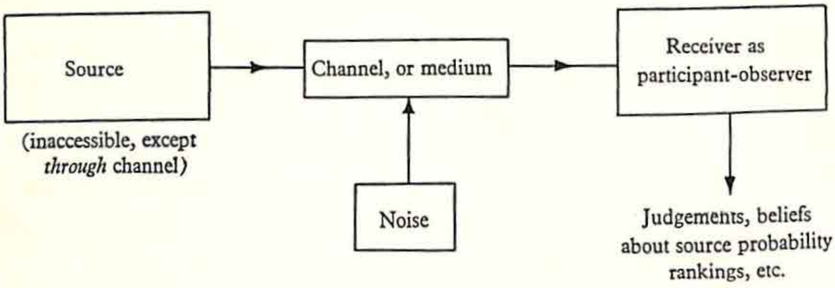

Source
(inaccessible, except *through* channel)
Channel, or medium
Noise
Receiver as participant-observer
Judgements, beliefs about source probability rankings, etc.

Fig. 2. The participant observer, or receiver as his own observer.

$$H(x) = -\epsilon(\log p(x)) = -\sum_x p(x) \log p(x),$$

$$H(x/y) = -\epsilon(\log p(x/y)) = -\sum_x \sum_y p(x, y) \log p(x/y),$$

$$R = H(x) - H(x/y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)\, p(y)}.$$

This last expression may of course be written $\epsilon\left((\log \frac{p(x/y)}{p(x)}\right)$ so that we have the valid concept of the 'rate of communication of information', between these two points, as the expected value of the logarithmic ratio of two probability distributions, the source distribution $p(x)$ and the conditional probability $p(x/y)$. Again, as is well known, this rate of information may be written another way:

$$R = H(y) - H(y/x) = \epsilon\left(\log \frac{p(y/x)}{p(y)}\right),$$

giving the concept of true information rate as the difference between two rates, that potentially existing in the received noisy signals, $H(y)$, some being bogus, and the rate of this bogus information, $H(y/x)$, arising from the noise. These rates are assessable only by the external observer for only he sees both ends of the channel and hence 'knows' x.

The point to be emphasized here is that this ratio R is described, in this way, entirely in terms of complete probability-*distributions* (averaged, at that); information is expressed by the conversion of a prior to a posterior *distribution*. No reference is made to the final process required in any physical, useful, receiving apparatus, such as a teleprinter—namely, the printing of some unique key-sign each time a signal is received (its 'overt behaviour', or reactions). The best the detector can do is compute the required probability (or likelihood) distribution; some threshold mechanism must determine the specific choices of x out of $x_1, x_2, \ldots, x_n$ at each signal-receipt, as communication proceeds. This choice then throws away information each time it is made.

Nevertheless, in a communication system, the observer can equally well observe the printed letters of the teleprinter. Some will fail to correspond with the letters at the source, and errors will occur; the observer can take these printed letters as being the 'received signals' y and so, regarding now the teleprinter apparatus as being part of the complete channel, compute the rate of information right through, from the source keyboard to the receiver keyboard. Thus the same measure can be applied at successive points along a communication system, all being accessible, and the rates of information calculated, up to each of these points, as required.

Now a word about the participant observer. By this name we would imply, when discussing communication systems, one who is situated at the receiver end, and having no access to the source except *through* the communication channel itself. He can assess the probabilities of the transmitted message-signals x only by inference from his noisy received signals; again he may have no access to the noise source, to observe it. Consequently, he cannot describe the whole communication process; communication theory, very properly, is not expressed in *his* language. He cannot, for example, compute the information rate R, nor the channel capacity C, because both involve x.

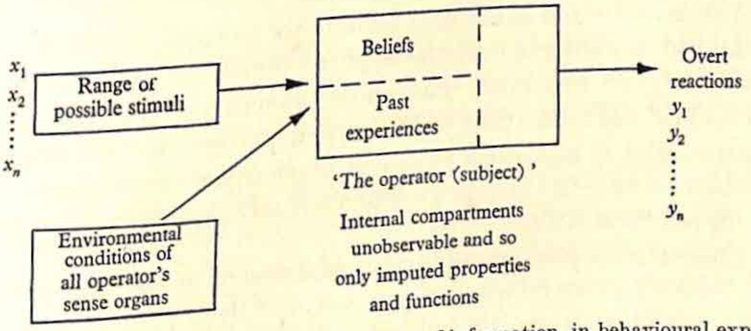

x_1
x_2
x_n
Range of possible stimuli
Environmental conditions of all operator's sense organs
Beliefs
Past experiences
'The operator (subject)'
Internal compartments unobservable and so only imputed properties and functions
Overt reactions
y_1
y_2
y_n

Fig. 3. The human operator (subject) and the transfer of information, in behavioural experiments.

Information transfer and decision making

In communication theory we discuss the transfer of information, and express it quantitatively, in terms of the ratio of a prior to a posterior probability distribution. Thus, if x represents a set of possible sent message-signals $(x_1, x_2, \ldots x_n)$ which become perturbed by noise, and if y_e is a unique noisy signal on some *one* occasion, then the receiver can be regarded as computing the likelihood function $L(x)$ where $L(x)=p(y_e/x)$. The *whole* of this function represents received information. This does not determine which particular x of the set $(x_1, x_2, \ldots, x_n)$ should be taken as the truth, and this *decision* depends upon some criterion of detection built into the receiver. The whole of this function contributes towards information about x; if one x be chosen as 'true', and the other possibilities rejected, then information is lost. Communication theory does not, as it stands at present, consider the bases on which such decisions may be made, though a whole theory is concerned with this very type of problem.

Let us consider a typical psychological experiment, in which an experimenter selects signals from a set of alternatives and applies them as stimuli, one at a time, to a human operator (subject). The environment is perhaps noisy, or disturbing in other ways, so that the operator is in the situation of having to identify the alternative signals selected, from these perturbed stimuli, perhaps whilst he is under stress. For example, the signals selected may be random words, the stimuli spoken utterance of them, while the environment may contain acoustic noise; the operator may then be required to guess the words from the sounds he hears.

It is convenient to think of the brain of the human operator as being a superb device designed to form and weigh hypotheses according to evidence supplied to it. Thus, with reference to Fig. 3, a stimulus received in noise or other environmental uncertainties constitutes evidence to the operator about the possible selected signals (alternative hypotheses); the operator may use this evidence and weigh his various alternative hypotheses (the likelihood function), finally deciding and making overt reaction, representing his final decision (not necessarily consciously). But from his overt behaviour alone we cannot discuss this decision process, at least in the language of Communication Theory, for this requires an external observer. Decisions are made in ways which depend not only upon the evidence and the operator's past experience, but perhaps upon his expectations of the future consequences of the decisions too. Decisions are taken in varied ways, and may depend upon maximizing chances of survival, or bodily comfort, or one's economic situation; in fact upon a range of possible 'utilities'.

An external observer (who is entitled, at least, to try and apply Communication Theory to his observations) can observe only the applied stimuli, the environmental conditions and the operator's overt responses. He cannot observe, or report upon, the subject's hypotheses, his likelihood estimations or his decision processes, at least separately. He can at best speculate. But it is possible to design psychological experiments in which constraints operate so as to define clearly a set of alternative hypotheses (as, for example, in Hick's experiments, 1952, 1953).

Although the complete likelihood function represented in the subject's brain cannot directly be observed, it may be taken into account descriptively, in a familiar way. Thus, before a stimulus arrives, the subject will have formed a number of hypotheses, each relatively weighted, concerning this stimulus (prior 'likelihood') representing his 'prior state of expectancy'. The receipt of this stimulus, in the particular disturbing environment, will normally change this weighting distribution, resulting in the subject's changed attitude, set, or 'posterior state of expectancy'. Depending upon whether the successive stimuli are related or not, the subject's 'state of expectancy' towards subsequent stimuli may be changed. Such 'states of expectancy' are, descriptively, like likelihood functions and might be correlated with actual functions (matrices) represented neurally in the brain. But in psychological experiment these are not normally observed. The subject's overt reaction to a stimulus is observable, and this is *assumed* to depend upon the 'prior state of expectancy' and the stimulus; but the precise nature of this dependency ('decision process') cannot readily be observed.

Nevertheless, it may be possible for an external observer to *infer* something about the likelihood function in the subject's brain (giving 'states of expectancy') by observing long series of overt reactions to successively applied stimuli.

IV. SOME EXAMPLES OF EXPERIMENTAL RESULTS VALIDLY INTERPRETED IN COMMUNICATION THEORY TERMS

I should like now to give a few illustrations from recent well-known experimental work, but I must emphasize that any criticism offered does not reflect upon the value of the experiments *per se*, or on their method; my comments refer solely to the use of Communication Theory and concepts in their interpretations.

(i) *The choice reaction-time experiments of Hick*

In these experiments, visual stimuli are derived from a number of small lamps, arranged in a circle, lit one at a time at 5 sec. intervals in virtually random order and with equal frequencies. The subject responds by touching one of a corresponding number of keys. It is found that the total time taken to respond is not a constant 'reaction-time', but increases as the logarithm of the number of lamps or keys in the set (with a certain correction at the low number end). These experiments are now familiar so that description may be left in such brief, fragmentary form.

In such experiments Hick showed a full appreciation of the assumptions of Communication Theory. Thus, it is reasonable to assume that one lamp is much like another, and so semantically neutral. In his earlier experiments, Hick operated his successive lamp-lighting at low speeds, so that zero errors occurred; there was then no question of 'noise', or of 'equivocation'. Further, he used long sequences and averaged the results; he also took the precaution, which we noted earlier, of averaging the results over an ensemble of operators. The results were observed by an independent 'external observer'. So all the necessary conditions applied.

In a second series of experiments, Hick required his subjects to shorten their reaction times, even at the expense of making errors. This situation corresponds to a noisy communication channel if, as before, x represents the transmitted signals (stimuli), and y the received ('noisy') signals (responses). Then R, the rate of information transfer is obtained from the formulae given above:

$$R = H(x) - H(x/y),$$

where $H(x)$ and $H(x/y)$ can be assessed from the frequencies $p(x)$, $p(x/y)$ of stimuli and responses over the set allowed.

Notice that here R is not the rate of information about x to the subject (operator). It is the rate of transference *through* him, as though he was a noisy or unreliable part of a communication link. This is a perfectly valid and useful concept, and it does not involve the 'likelihood estimations' and 'decision process'.

Again, no attempt has been made to give exact interpretation of the operator's 'capacity' in the technical Communication Theory sense of that word; for 'capacity' involves the concept of the hypothetical maximum information rate, over all possible forms of coding of the stimuli (Shannon, 1949). And we cannot try all possible forms. However these experiments suggest that R might be used as a valuable measure of the degree of matching between an information source and a human operator.

(ii) *Application to intelligence tests*

By contrast, I do not find myself in such ready sympathy with Hick (1951) in his uses of Communication Theory concepts with reference to intelligence tests. In his paper, he first discusses deduction and deduction-induction as the possible processes which may be applied to a test question, in order to frame an answer, with the inevitable probabilistic nature of any answer on the latter basis. I would agree with the author about the hollowness of arguments raised against induction by some theorists; induction is all we have (besides which, equally weighty arguments can be applied against the concept of 'deduction', e.g., see J. S. Mill's *Logic*). Hick then draws analogy to the extraction of messages from a noisy communication channel; the 'questions' used in intelligence tests represent 'noisy signals' and the 'answers' represent 'extracted messages'. The operator, whose intelligence is being tested, is then 'a predictive filter'.

Functionally speaking, this comparison is interesting, but it is the next steps with which I incline to disagree. 'Noise' is compared to the 'agency of the psychologist, who frames the questions', and 'is probabilistic *from the subject's* (*operator's*) *point of view*' (italics mine). But in Communication Theory the receiver has not got a point of view; it is written wholly in the language of the external observer. Further, Hick discusses the probabilities of rules as applying to questions as they are *to the subject*. Again, such probabilities, as beliefs, are non-numerical and do not come into the Theory of Communication. We may certainly discuss 'information' qualitatively, in such terms, but not numerically.

To repeat, I am not questioning the validity of Hick's results or theory; I simply wish to stress that it goes beyond established Communication Theory as this is used for numerical *measurement*.

(iii) *The construction of statistical approximations to language*

In my opinion, one of the most valuable additions to method in experimental psychology has come, from Communication Theory, with the suggestion of 'probability texts'.

In a human language, such as English, a great degree of redundancy arises from the existence of syntactic rules, which are obeyed by its users; such rules are variously described as rules of spelling, rules of grammar, etc. Of course, the user himself may not know the rules, any more than he knows *why* he laughs at a joke. Rather it is the students of language, the linguist and grammarian, who describe the rules, as they discover them, the user merely conforms to *habits*.

Rather than describe the rules in the sophisticated way of modern grammarians, Communication Theory gives means of assessing the constraints upon printed language in terms of sets of conditional probabilities. From counting great seas of printed texts we can assess frequencies of monograms, di-grams...n-grams, letter-bridges, etc., and describe the rules statistically. Unfortunately, it is questionable whether similar statistics could ever be found and prescribed usefully for spoken language. For speech is not spoken writing; it is unpremeditated and freely breaks all the rules; it is fragmentary and ungrammatical; and all conversations are specialized and topical.

Statistics of many printed languages have been assessed, but are reliable only up to tri-grams at the very most. Using these tables, sample tests may be constructed which approximate to the language. For closer, n-gram, approximations a working technique may be employed which is due to Shannon (1951), and which may be used for constructing both letter sequences and word sequences: a sequence of n words, chosen from some existing text, is shown to a member of the language group, who is asked to guess the next word; then the first word of this $(n+1)$ sequence is covered up and the new sequence of n shown to another member, who in turn guesses a suitable next word...and so on. Very large numbers of people from the same culture are required for constructing such 'probability texts' of any length to be representative of the language, on the assumption that they conform to their engrained language-habits, on an average.

Such 'probability texts' read as nonsense, but the degree of nonsense is controlled and is assessable *quantitatively*. The use of letters, words (nonsense or real), and sentences or long texts, as stimuli in psychological experiments is of course very old. But in the past it has largely been left to the psychologist to 'invent' or 'design' such texts to what he considered various standards of sense, meaningfulness, familiarity,..., according as his experience had taught him. But using this probability technique we can construct texts which are, say 10% or 50% like English, on an average. Their uses for experiments are endless (Miller, 1951).

To refer back to Sect. IIId; 'meaning' is essentially a *relationship* between a stimulus and its recipient. We are not denying that 'meaning' and 'meaningfulness' are not perfectly good, useful, concepts. Rather, it seems that the use of 'probability texts' by-passes philosophical discussion and substitutes a quantitative measure to the stimulus-recipient relationship, averaged over a population of recipients, and expressible in the meta-language of an external observer. For instance, reaction times may be found to vary according to the 'meaningfulness' of word or sentence stimuli; but, alternatively, ensemble-average reaction times may be related to the probabilities of these stimuli, quantitatively.

Tachistoscopic experiments

It has been known, at least since 1885, that speed of visual perception, as exampled by reading, increases with familiarity with the language. Using these 'probability texts', the exact functional relationship can be established. Employing tachistoscopic presentation Miller, Bruner and Postman (see Miller, 1951, for various applications to experimental psychology, including this experiment) measured the degree of correct identification of 8-letter 'words' having respectively random order, and first-, second- and fourth-order transition-probability approximation to English, when the visual exposure times were varied. Table 1 illustrates some of their test-'words'; from left to right these becoming progressively and measurably closer to English, on an average.

Table 1

Random order (zero redundancy)	1-gram (15% redundancy)	2-gram (29% redundancy)	4-gram (44% redundancy)
YRULPZOC	STANUGOP	WALLYOFF	RICANING
OZHGPMTJ	VTYEHULO	THERARES	VERNALIT
DLEGQMNW	EINOAASE	CHEVADNE	MOSSIANT
GFUJXZAQ	IYDEWAKN	NERMBLIM	BITTLERS
WXPAUJOS	RPITCQET	ONESTEVA	ONETICUL
etc.	etc.	etc.	etc.

Miller, Bruner and Postman have shown that although the 4-gram 'words' are recognized more quickly and accurately, in tachistoscopic tests, than those of poorer approximation to English, the rate of information intake (or rather *transfer*) is roughly constant. The 4-gram 'words' are recognized more quickly, but they are also more 'redundant'.

Provided that the results of such tachistoscopic tests are averaged over long series of stimulus words and, again, averaged over many subjects who are within the language group, such uses of the Wiener-Shannon information measure are perfectly valid, and very useful. Qualitative results become quantitative.

This illustration provides a good example of non-stationary conditions. Thus we cannot validly take time-averages because, if the tachistoscopic exposure is long enough, *every* letter of a word will be recognized without error. But averaging may be done instead over an ensemble of people (ensemble average).

'Probability texts' may be used for various quantitative psychological experiments upon visual perception and discrimination. For verbal association tests, for memory tests (recall, or learning), and for articulation tests under noisy conditions, for example. The writer has suggested elsewhere their use for assessment of the perceptual rates of visual information whilst reading—the relation between the frequency and distribution of eye saccades and the probability of the textual matter (Cherry, 1955).

These varied examples of psychological experiments which I have cited have been very few in number, but may serve to illustrate the points which should be borne in mind when interpreting results quantitatively with the statistical 'information-rate' measure.

V. INFORMATION RATES ARE RELATIVE TO A GIVEN SOURCE ENSEMBLE

The ring of N lamps used by Hick (Sect. IV (i)) and the 'probability texts' just described represent two ways of excluding 'meaningfulness' from experiments and thereby causing serious personal variations of response. In Hick's case, one lamp is as good as another, semantically almost neutral; in the 'probability texts', each dummy word may certainly be meaningful to any one subject, in the sense that it will set up images and associations in his mind, but the trouble is presumed to be overcome by the smoothing which takes place when results are averaged over many subject's responses.

These two kinds of stimuli represent different types of source ensemble. Hick's lamps form an alphabet of N members, having equal or other assigned frequencies; the words constructed for 'probability texts' may also be used in alphabets of N words, but their probabilities will, subjectively speaking and only on an average, conform to those of letters in 'the English language' represented by the habits of the subjects who are drawn from an English population. (Strictly, we should not speak of '*the* English language'; we might, for some purposes, distinguish journalese, scientific English, nineteenth-century novels, and a host of other languages. But a great number of statistical counts have been published, e.g., see Zipf, 1949.)

It is then important to appreciate that information-rates are always determined relative to some given source ensemble; an alphabet, with its set of probabilities. Information rates are not absolute, but relative to some frame of reference, in any specific practical case. When information rates are quoted, the ensemble should be specified, that is, in relation to psychological experiments, the exact stimulus conditions must be noted.

I have mentioned particularly visual and aural stimuli but, with this precaution in mind there is no reason why reliable and quantitative rates should not be found for other sense-stimuli, taste or smell for instance (Beebe-Center, Rogers & O'Connell, 1955). One can always set up an empirical alphabet of smells or tastes and measure information-rates *relative* to that frame of reference.

VI. THE UNREALITY OF 'CONTINUOUS INFORMATION'

Many tracking experiments may appear at first to use a 'continuous' stimulus function; for example, when sighting a gun on to a moving target, or when following a wandering line on a moving band of paper with the finger. Such notions raise the concept of 'continuous information'—a concept which to my mind has no place in physical experiments—and I should like to end by commenting upon it. But first, I would remark that 'continuous information' may well be a useful *mathematical* concept, for mathematicians (pure) deal with constructs, not with experiments and observations.

As is well known the information rate of a hypothetical noiseless source $H(x)$ can be infinite; 'continuous information' can be discussed meaningfully only if noise is considered, so that the equivocation $H(x/y)$ is infinite to the same degree. Their difference $R = H(x) - H(x/y)$ is however finite, so that everyone is satisfied. But in tracking experiments it may be difficult and undesirable to specify or account for any noise. The signal to be tracked is 'continuous' in a 'noise-free' environment. Then the question arises: is the rate of information supplied to the operator infinite?

This question is not answered by observing that the operator's perceptions and reactions must be discrete and finite in number. For instance, it might be pointed out that the operator's eyes do not track continuously, but in discrete saccades, his perceptions taking place during the fixation pauses only, when his eyes are temporarily at rest. Indeed, it would seem to me that 'perceptions' are of such a nature that they *can* only occur at points in time. The decision is made, the working hypothesis selected, at some instant; perceptions cannot be 'continuous', by their very nature.

Again, all human sense-organs have a minimum threshold, below which they are not activated; finite changes in stimulus are needed to set up perceptions. However, these objections do not answer our question, because they refer not to the source of information, but to the recipient's perceptions and reactions.

If the question raised here be considered proper to Communication Theory, the answer to it is *no*. For it is the external observer who gives the answer, not the operator himself, and when the observer describes anything he can say only a finite number of things about it. If the 'continuous' curve being tracked consists of a wavy line on paper, the observer can state the position of a finite number of sample points to a finite accuracy. But he cannot describe the whole curve. It may be objected that he could quote the equation of the curve; but still he is giving only an equation, symbolically, not a curve drawn on paper. Again, he cannot specify the tracking subject's responses, his moving finger positions for example, with mathematical precision.

In practice, the information rate of a physical 'continuous' source, as may be used in tracking experiments, could be specified by choosing some agreed set of spaced samples and dealing in discrete terms, with the amplitudes of these samples quantized. The stimuli and responses are reduced to a finite set, with a finite number of degrees of freedom. The information rate of the source, and also of the operator's reactions (output rate) *are specifiable only with respect to some prior agreed discrete set of choices.*

But the concept of a 'continuous information' source certainly has its uses, and perhaps a distinction may be drawn which clarifies the point.

Pure and applied Communication Theory

Like other branches of mathematics, Communication Theory has its 'pure' side, to which the concept of a 'continuous' source rightly belongs. Continuity is essentially a mathematical concept and, of course, a most valuable simplifying concept. It has been easier, quicker and more fruitful to the development of mathematics to deal with continuous functions than with great masses of finite data.

Again, there are those whose prime interest is in Communication Theory as a mathematical discipline, and not in its application to the description of physical communication situations. Whereas the physical situations call for an observer, in whose meta-language the description is made, mathematics calls for no observer at all—no meta-channel of description.

REFERENCES

BEEBE-CENTER, J. G., ROGERS, M. S. & O'CONNELL, D. N. (1955). Transmission of information about sucrose and saline solutions through the sense of taste. *J. Psychol.* **39**, 157.

CHERRY, E. C. (1955). Information, feedback and the human senses. *Trans. Soc. Inst. Tech.* **7**, 81.

HICK, W. E. (1951). Information theory and intelligence tests. *Brit. J. Psychol.* (*Stat. Sec.*), **4**, 157–64.

HICK, W. E. (1952). On the rate of gain of information. *Quart. J. Exp. Psychol.* **4**, 11–26.

HICK, W. E. (1953). Information theory in psychology. *Trans. Institute of Radio Engineers, Professional Group on Information Theory.*

MACKAY, D. M., (1950). Quantal aspects of scientific information. *Phil. Mag.* **41**, 289.

MILLER, G. A. (1951). *Language and Communication.* New York: McGraw-Hill.

SHANNON, C. E. (1949). *The Mathematical Theory of Communication.* Urbana: University of Illinois.

SHANNON, C. E. (1951). Prediction and entropy of printed English. *Bell Syst. Tech. J.* **30**, 50.

ZIPF, G. H. (1949). *Human Behaviour and the Principle of Least Effort.* Cambridge, Mass.: Addison-Wesley.

(*Manuscript received* 10 *May* 1956)

ON SOME PSYCHOPHYSICAL TECHNIQUES EMPLOYING A SINGLE MANIPULANDUM

By LAWRENCE WEISKRANTZ

*Laboratory of Physiology, University of Oxford**

Four techniques are described which permit the single bar-press apparatus to be used for the generation of psychophysical functions. Rate of response is a useful dependent variable for measuring the strength of behaviour associated with the negative stimulus in two of the methods. When a warning signal is introduced with discrete negative and positive stimuli, the situation becomes logically akin to the traditional infra-human psychophysical procedures. Rate of response is then no longer useful as a variable, but because any given response does not alter the animal's geographical position, the bar-pressing situation offers certain advantages over the traditional procedures. These advantages are incorporated in the two remaining techniques described.

I. Introduction

In recent years there has been an increase in the use of automatically controlled 'vending machine' types of apparatus, in which manipulation of a single lever, bar, or panel can be made to deliver food, or other reward, according to conditions set for the experiment. (For an excellent description and analysis of the single manipulandum situation, see Ferster, 1953.) Most commonly, these devices have been employed for the study of behaviour under conditions where there are no environmental stimuli which are distinctive occasions for rewarded responses—such as the study of rates of response under variable or fixed intervals schedules of reward (Skinner, 1953). Occasionally, they have been used for the investigation or demonstration of discriminated behaviour (Skinner, 1953; Heise, 1953). The latter application is likely to increase in the future, if for no other reason than that bar-pressing units are approaching the status of standard equipment in many laboratories, thereby tempting research students by their availability, and because the units have been made automatic to such an extent that considerable benefits are offered in economy of time and flexible control of stimulus scheduling. It is worth inquiring, therefore, into the relationship between the bar-press, or single manipulandum situation, and procedures traditionally employed for the study of psychophysical problems. Such an inquiry should yield information of both practical and methodological value.

For practical purposes, the single manipulandum situation is limited to successive discriminations, although simultaneous discriminations are not beyond the range of possibility. Its traditional counterparts for infra-human subjects, therefore, are such situations as the Y-maze, the single-stimulus Lashley jumping stand, and the Wisconsin discrimination apparatus when arranged for a 'go—no-go' or approach-avoidance discrimination. The bar-press situation, however, can differ in two respects from these other situations. In the first place, each unrewarded response leaves the animal in exactly the same physical situation as before the response, and hence rate of response can be used as a dependent variable. In the second place, in the traditional situations the stimulus permits the animal to choose one of two alternative responses, i.e. *where* to respond, while in the bar-press situation, the stimulus may only tell the animal *when* to respond. But, as we shall see, the bar-press situation also can be made logically equivalent to traditional situations.

---

* Now at The Psychologicol Laboratory, University of Cambridge.

The limitations of a given procedure are more evident when one considers psychophysical functions rather than simple discriminations. There is little difficulty in establishing a 'discrimination' between, say, two widely separated acoustic tones, if one tone is always made the occasion for a given schedule of reward to be operative, while the other is never the occasion for a reward. Indeed, this serves commonly as a class-room demonstration. But it would be impossible to determine the *DL* (i.e. least discriminable difference) between two tones with these contingencies, at least not without imposing rather strict limitations on such factors as stimulus duration relative to average reward interval. Consider the instance of a stimulus difference which is subliminal, or in the limiting case, zero. Given even a low 'operant' level of response rate, the animal would be rewarded occasionally for responding to a stimulus difference which was undetectable; when this occurred, it would be to the animal's advantage to continue responding. The discrimination could be based on the occurrence either of a reward *or* of a supraliminal stimulus difference. Put in other terms, the difficulty lies in the non-independence of rewarded responses over a period of time; given one rewarded response, the probability is high that a response during a succeeding given interval will also be rewarded, quite independently of the presence or absence of any other discriminative stimuli. One step towards avoiding this difficulty is to make the reward terminate the positive stimulus. All of the methods discussed here will be subjected to this condition. But it is evident from what has been said that psychophysical requirements refine the issues to be raised in evaluating a discriminative situation. All of the illustrative material here, therefore, will be of psychophysical determinations.

All successive discriminations share one difficulty. In the early stages of discrimination training, there is usually a high degree of 'generalization' or 'induction' between the two stimuli; i.e. the animal makes the same response to both stimuli. In a 'go—no-go' discrimination, for example, where the animal is required to make a response to one stimulus (the 'positive' stimulus) and to refrain from making that response to the other stimulus (the 'negative' stimulus), the animal will sometimes continue to respond to both stimuli. Since indiscriminate responding produces intermittent reward, and since intermittent reward schedules are known to produce relatively high and stable rates of response, the pattern may be an exceedingly persistent one. The real problem, then, often lies in first reducing the probability of an incorrect response to a negative stimulus rather than increasing the probability of a correct response to a positive stimulus. One of the traditional methods of achieving this is to punish incorrect responses. The use of aversive stimulation, however, involves many additional difficulties, such as the suppression of all behaviour in the situation. Fortunately, other procedures are available which do not employ aversive stimulation. Some of these will be discussed below. They can be placed in one or both of two classes: (1) extinguishing the incorrect response to the negative stimulus—the negative stimulus is presented either much more frequently or for longer duration than the positive stimulus; (2) making the appearance of the positive stimulus contingent on the subject's not having made an incorrect response to a negative stimulus. Both classes of procedures have several variants, each of which has special problems—but these are problems which are basically inherent in the actual meaning of discrimination. In this sense, it is a truism that there is no available technique which is actually, or potentially, devoid of such problems, any more than there is a technique which is not a function of variables other than the stimulus dimension.

This report will discuss four specific psychophysical procedures which have been evolved for use in the bar-pressing situation. The first two find no parallel among the traditional procedures, in that they employ rate of response as a dependent variable. The last two have traditional counterparts, but the bar-press situation offers certain additional advantages. All illustrative material will be drawn from two types of threshold studies: the difference limen for frequency, using acoustic stimuli; and the critical fusion frequency, using optic stimuli. Subjects include pigeons, rhesus monkeys and humans.

II. CONTINUOUSLY CHANGING STIMULUS METHOD

If the positive stimulus is terminated by reward, it can be asked whether rate of response has any value as a dependent variable—a single response to the positive stimulus can yield as much information to the experimenter (and to the subject) as N responses over a time interval. But there is still the negative stimulus, and here rate of response might be used effectively. One ingenious possibility has been reported by Skinner (1953), who studies the behaviour of pigeons with respect to what he termed an 'external clock'. A straight line on the front panel of the cage was made to increase in length at a constant rate, the bird being rewarded for responding only when the line had reached a certain maximal length. After the reinforcement, the stimulus reverted to its original value and went through another cycle. The characteristic behaviour which finally emerged was a lack of bar-pressing until just before the stimulus had reached its final value, when the bird responded at a high and relatively fixed rate until rewarded. The *DL* can be defined as the value at which the animal begins to respond, the mid-point between this value and the final reward value, or some other fixed value.

Two stimulus parameters can be assigned. The first is the stimulus difference between the starting value and the final value; the second is the rate at which the change of the stimulus occurs—or, in more general terms, the form of the transition function from starting to final value. These factors determine, of course, the temporal length of the cycle from start to finish. As yet, no completely systematic investigations have been made of these parameters, but there is evidence which suggests that variations in these factors can produce large differences in behaviour, and that the relationship is probably not a simple one.

Skinner, after comparing cycle periods varying from 3 to 32 min., states that 'the rate of responding is roughly the same for a given size of spot regardless of speed of change. The curve at 32 min. is obviously not approximately 10 times as high as that at 3 min., as it should be if the control by the spot were strictly equivalent' (p. 74). This means that the *DL* became smaller as the temporal period increased. In the published curves (which are unfortunately too small to permit precise measurement), however, there is a suggestion of a rather smoother and more gradual start with long periods than with short periods, and the starting-points, in terms of stimulus value, might be roughly the same.

Using shorter periods of 1 or 2 min., the behaviour can be shown to be affected by the stimulus difference between the starting and final values. A tone was made to warble (at the rate of three times per second) between a value of 2600 c.p.s. and another variable frequency. The second frequency was gradually and continuously lowered until the tone remained steady at 2600 c.p.s.; the pigeon was rewarded for responding to the steady tone. The entire run took 56 sec., the steady tone remaining on for the last 5 sec. of the run. After the steady tone, there was a 5 sec. period of silence, and a new run was begun. The initial warble value could be varied without changing the decay function. In fig. 1 are shown three such values, plotted against rate of response. Each point is actually an average of 15 runs. It can be seen that as the starting value is lowered, the animal begins to respond earlier in the run. But the point at which the animal responds is not solely under the control of the absolute stimulus value. If mid-points are chosen as the *DL*'s, the values are 55, 30, and 11 c.p.s., respectively. Substantially the same results were obtained with a run of 2 min., although the data are not as complete for that value.

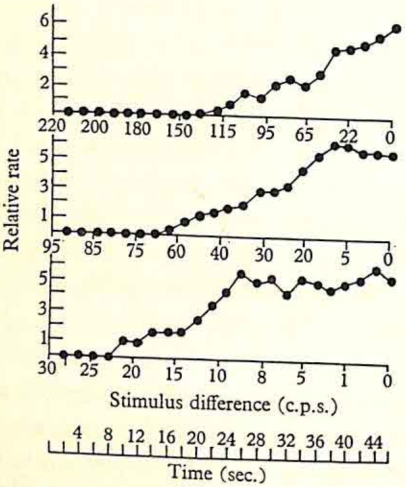


Fig. 1

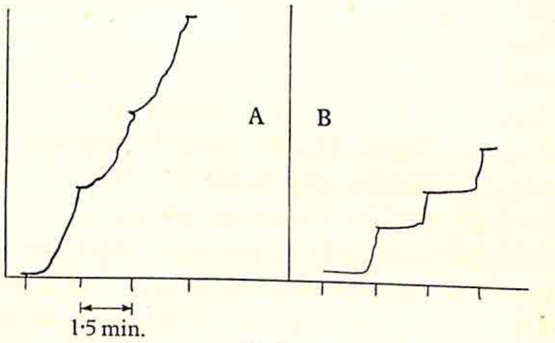


Fig. 2

Fig. 1. Average relative rate of response for a pigeon as warbling tone approaches steady tone. Three curves are for three different initial warble ranges.

Fig. 2. Successive cumulative response curves for a pigeon as warbling tone approaches steady tone. (A) Rate of stimulus change is independent of the animal's response. (B) After 2½ hr. training in which 'early starts' locked the stimulus until animal stopped responding.

It is quite likely that if longer runs were chosen, the behaviour would tend to become less and less a function of the initial stimulus difference, but, at least within the range explored, the initial value can influence the relationship considerably. It will also be noted that the curves in Fig. 1 show a fairly gradual increase in rate as the stimulus approaches the steady tone. This was not produced merely by virtue of the averaging of values. The response pattern, typically, showed a 'scallop' rather than a 'staircase', effect. Fig. 2 A shows a tracing of three successive runs made by the pigeon immediately after he had yielded the data for Fig. 1. Again, such scalloping might tend to sharpen up with longer runs or with increased training—the present curves were the product of 14 days of training, 2 hr. daily.

It is possible, however, to sharpen up the curves by means of one of the general methods discussed in the Introduction. When human subjects were trained on this discrimination (see below), a common verbalization was 'the tone was still warbling but I thought I might as well start early and make sure'. There was no contingency, in other words, forcing a low rate of response during the negative stimulus. In order to eliminate early starts, an arbitrary stimulus can be chosen such that, if the animal begins to respond at a larger value, the stimulus is instantly stopped and kept at that value until after responding has stopped. Early starting thereby delays the appearance of the steady tone, and therefore of reinforcement. That this technique was effective in sharpening up the behaviour is clear from the cumulative tracings shown in Fig. 2. In this experiment a run of $1\frac{2}{3}$ min. was used, including a 10 sec. silence period. Fig. 2A represents the responding at the beginning of the experimental session, without the added 'stimulus locking' contingency. Fig. 2B shows the same animal $2\frac{1}{2}$ hr. later, after the contingency had been introduced, but in these particular trials it had been removed, and conditions were otherwise the same as in Fig. 2A.

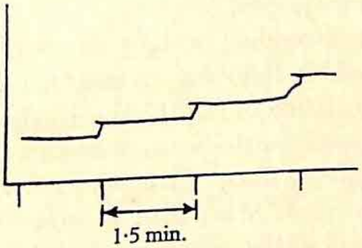


Fig. 3. Successive cumulative response curves for a human subject as warbling tone approaches steady tone, using the 'stimulus lock' device described in text.

Human subjects can be trained to make a similar discrimination. The subject is provided with a handswitch and a set of earphones. His only instruction is that he is to make a light on a panel flash on as often as possible. Roughly the same training procedure is then employed, including the addition of the 'locking' of the stimulus for early responding. Shown in Fig. 3 is a cumulative response curve, using a warbling tone which approaches 2600 c.p.s. This subject started to press at about 8 c.p.s. difference from the steady tone. Training time has been found to vary among subjects from as little as 30 min. to as long as 6 hr.

In general, it is felt that this method has not yet reached a point of refinement where it provides a useful measure of discriminative capacity. Under certain conditions, it is extremely sensitive to general parametric conditions and, in addition, is probably very dependent upon length and method of training adopted. These are not 'faults' of the methods; all methods are affected similarly, but, in the opinion of the author, this particular technique is more sensitive to these factors than are many other procedures. There is an additional difficulty in general method—this will receive discussion in a later section. But, if for no other reason than the speed with which psychophysical functions can be generated, the method deserves further investigation.

III. BACKGROUND EXTINCTION METHOD

In the previous method, the stimulus 'locking' device was of the second type described in the Introduction for reducing the frequency of incorrect responses to negative stimuli—making the appearance of the positive stimulus contingent on the subject's not having made a bar-press to the negative stimulus. The other technique—extinction—is well represented in the present method. As before, rate of response is used as a dependent variable with respect to the negative stimulus, but an additional dependent variable is introduced in the simple percentage of responses to the positive stimulus. The psychophysical function is calculated as a joint function of both variables.

The negative stimulus is presented continuously as a background against which short durations of the positive stimulus are presented on an aperiodic schedule. For example, in the determination of the absolute threshold for acoustic intensity, long periods of silence are presented between short presentations of tones of varying intensity. Heise has used such a technique quite successfully for determinations of the minimum audible field of the pigeon (Heise, 1953). In another example, in the determination of critical fusion frequency, long periods of steady light are presented between short presentations of flickering lights of varying frequency.

There will always be some responding in the presence of the background or negative stimulus. Such responding, which Heise has termed 'noise' if it is of significantly high rate, must enter into the calculation of the threshold value, since it can provide a high 'chance' probability of a response to the positive stimulus. A simple mathematical correction for noise is available (Heise, 1953). The most effective policy, of course, is to lower the noise level to a negligible value, rather than to correct for a high level. One method is to increase the mean interval of the aperiodic schedule. For a pigeon under ordinary conditions of deprivation, a schedule with a mean of 2 min. has been found quite satisfactory for the auditory discrimination just described. An additional possibility, one used by Heise, is to increase the number of responses required for reinforcement during the appearance of the positive stimulus. For example, one can require five bar-presses in 5 sec. This, in effect, redefines the unit response.

Regardless of these controls, noise level tends to increase to a considerable value when the positive stimulus is maintained at just supra-threshold values. For this reason, it is important to ensure that very long sequences of difficult values are not presented. Using this technique, therefore, to obtain thresholds by the 'quantal method' (Stevens, Morgan & Volkman, 1941) for infra-human organisms is not always feasible, although it seems at first glance to be almost ideal. More practical procedures are to present values either randomly or in short sequences of a given value.

Another factor to be taken into consideration is the prolonged presence of the negative stimulus. If the problem is one in which fatigue or adaptation is likely to be a factor, this method may have serious drawbacks. In a later section, inferential evidence will be presented suggesting that such an effect can be significant under certain conditions in auditory discrimination.

Fig. 4 shows a psychophysical function for acoustic frequency for a pigeon, obtained with the Background Extinction Method. The negative background stimulus was a steady tone of 2600 c.p.s., 60 db. sensation level (human), while the positive stimulus was a frequency-modulated tone which warbled at the rate of three times a second between 2600 c.p.s. and another variable higher frequency; as the variable frequency approached 2600 c.p.s., it became more difficult to distinguish a warble from a steady tone. The variable frequency is plotted on the abscissa; the ordinate represents the percentage of stimulus presentations to which there was a response of at least 10 bar-presses. The positive stimuli were presented on a 2 min. variable interval schedule, each stimulus having a duration of 5 sec. The positive stimulus value was kept constant for four or five presentations, and then changed to a new value randomly. Each point represents at least eight presentations. The *DL* (50% point) for this bird was approximately 25 c.p.s. No correction was made for noise, since it was found to be negligible.

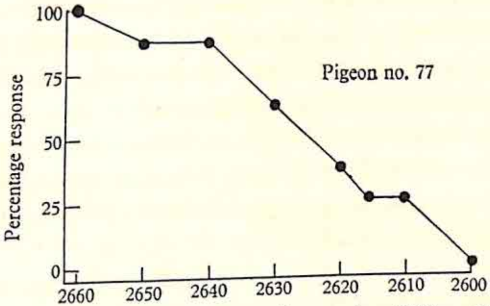


Fig. 4. Pigeon's psychometric function for acoustical frequency, obtained with the Background Extinction Method.

IV. METHOD OF LIMITS

Each of the previous methods has at least one difficulty associated with it, methodological in one case, technical in the other, which has not yet received attention. The Continuously Changing Stimulus Method evidently is insensitive without a stimulus-locking device, as described, since it is never to the animal's disadvantage to begin its response run early. But the locking device introduces a delicate problem of the actual definition of the *DL*.

The Background Extinction Method suffers from no such vagueness about the definition of the threshold but, when used for discriminations made along certain stimulus dimensions, it presents a serious technical problem. For example, consider the case of the *DL* for acoustic frequency. In changing from the background of frequency A to the positive stimulus of frequency *B*, acoustic transients will be produced—varying as a function of the actual transition curve. These transients may be audible as a click. A related difficulty may be created by the switching mechanism; there may be a brief period of non-stimulation or of summation of both stimuli. With carefully designed apparatus, these problems can be eliminated—but often only after considerable investment of money or time. In general, the Background Extinction Method, for these reasons, is more suitable for absolute rather than for relative thresholds.

An alternative approach is one in which the artifact is deliberately retained. On some occasions, however, the negative stimulus is substituted for the positive, a device sometimes called the 'introduction of blanks'. It is of interest to note that this simple alteration provides the bridge between the bar-press situation as it has been used heretofore and the traditional situations. In the latter, the animal is always given a cue or warning signal for the *opportunity* for response—it may be the raising of a screen or the placing of the animal in the starting box. Concomitantly, this alteration renders rate of response superfluous as a variable. Instead, responses to positive stimuli are compared with responses to blanks.

The following method is one which attempts to meet both of the methodological and technical difficulties just discussed. It bears its closest relationship to the method of limits as used in human psychophysics. The method of limits, without modification, is clearly unsuitable for animal work, since the reward contingency cannot be established satisfactorily—one cannot reward a 'correct' response. To illustrate the method used here, an example will be given of the determination of the *DL* for acoustic frequency, employing a warbling stimulus as described above.

Discrete tones of 5 sec. duration were presented. The first tone presented always had a warble range of fairly high value (i.e. greater than 100 c.p.s.). If the bird did *not* respond to the tone, the next tone was a lower value of the warble range. Similarly, if it did not respond to this, the next tone was a still lower warble. The animal was rewarded if it responded when the tone finally was steady, as it *might* have been after 4 or 5 steps. If the bird *did* respond to a warbling tone, regardless of its range value, the duration of the tone was extended for an additional 5 sec., during which time it was changed quite quickly to a much higher value of warble range. Then on successive trials the animal would be forced to start over again from this higher value, working down to lower values by not responding. In other words, any 'mistake' would necessitate starting the series of descending steps over again, thereby delaying the possibility for reward.

On the basis of how well the animal did in the earlier stages of training, a 'critical area' of warble range was selected. It was in this area that the animal no longer consistently discriminated the warbling tone from a steady tone (i.e. he sometimes responded to the warble). The critical area included two warble values, an upper and a lower. From a table of random numbers, two schedules were prepared, one for the upper and one for the lower value. If the animal reached the upper value in the critical area, a steady tone was presented 25% of the time instead of the usual warble range value. If he then responded, he was rewarded. If he did not respond to the steady tone, the series was begun again. A similar procedure was adopted for the lower value, except that on 75% of the occasions a steady tone was substituted for the warble. Hence, in effect, the schedule was one which tended to improve the discrimination, since the lower value was more heavily weighted. With such improvement, the animal would refrain more frequently from making incorrect responses to the lower warble value, in which case a steady tone was presented as the next stimulus. If the animal correctly refrained from responding to more than half of the lower warble presentations (out of an arbitrary number of presentations) the entire critical area was lowered. Similarly, if the animal was incorrect on more than half of the upper warble presentations, the critical area was raised. The *DL* with this procedure is defined in terms of the point at which the animal mistakenly responds to a negative stimulus.

The threshold values which have been determined in this fashion possess a high degree of stability. It is usually possible to run a pigeon for at least 2 hr., obtaining approximately 100 successive threshold values without any apparent large shifts in values. One principal advantage of the method lies in the speed with which successive *DL*'s may be obtained. For the data here discussed, a value was obtained, on the average, once in a period just longer than a minute. For this reason it lends itself to the study of short-term changes in thresholds. For example, Deutsch (1951) has reported a temporary decrease in the human monaural *DL* for pitch following stimulation of the contralateral ear for 4 min. by a tone of the frequency from which the *DL* is being measured. This effect he termed an 'auditory figural after-effect'. A similar phenomenon (with binaural stimulation) in the pigeon is suggested by Fig. 5, which shows successive *DL*'s for pitch at 2600 c.p.s. for a pigeon before and after a 4 min. presentation of a 2600 c.p.s. tone.

(When the 4 min. tone was presented, the overhead light was turned off, which effectively stopped the pigeon's responding.) In Fig. 6 is shown similar data for another pigeon, together with control data for the effect of a 4 min. 'dead time' period. In both graphs, there is a suggestion of a lowered *DL* following the 4 min. tone, with a gradual reversal back to its normal level after some minutes. (The *DL*'s shown in Figs. 4 and 5 were obtained from the same pigeon, no. 77. It is interesting to note that the *DL* following 'satiation' in Fig. 5 is roughly the same as that shown in Fig. 4. The latter was obtained by the Background Extinction Method, in which there was a steady tone present continuously, except for short bursts of the positive warbling stimuli. It may be argued, therefore, that both figures show the effects of a satiation tone on the *DL*.) Repetition of the Deutsch experiment using human subjects and the present technique, however, suggests that the effect is found only under rather special conditions, and the present data are certainly not complete enough to be taken as positive confirmation of his findings. They are presented, rather, to illustrate the use to which the method may be put.

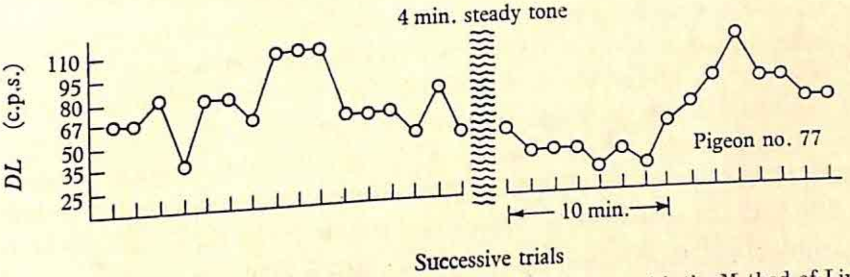


Fig. 5. Successive *DL*'s for pitch from a tone of 2600 c.p.s., obtained with the Method of Limits, before and after a 4 min. satiation tone of 2600 c.p.s.

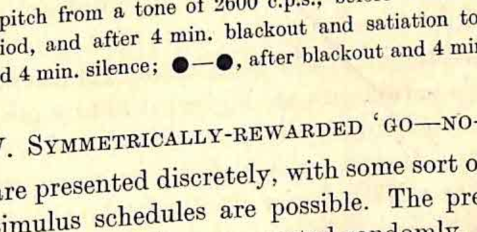


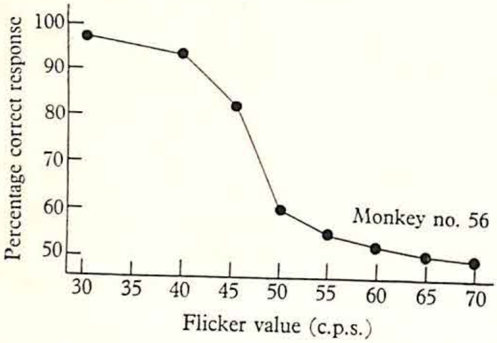
Fig. 6. Successive *DL*'s for pitch from a tone of 2600 c.p.s., before 4 min. blackout period, after 4 min. blackout and silence period, and after 4 min. blackout and satiation tone. ○—○, Before blackout; ———, after blackout and 4 min. silence; ●—●, after blackout and 4 min. steady tone.

V. SYMMETRICALLY-REWARDED 'GO—NO-GO'

So long as both stimuli are presented discretely, with some sort of accompanying warning signal, a great many stimulus schedules are possible. The present method is perhaps the most general, in that both stimuli are presented randomly. It is the analogue of the human 'method of constants'. This method has the advantage over the Method of Limits of permitting almost complete automaticity of procedure, and, concomitantly, in removing the experimenter as a controlling and possibly biasing factor. It loses, however, the advantage of speed in threshold determination.

The basic contingencies are simple. A stimulus is presented every 10 sec., say, with a maximal duration of 5 sec. If the stimulus is positive, the animal is rewarded for a bar-press.

If the stimulus is negative, the animal is rewarded at the end of the 5 sec. period for *not* responding. If the animal behaves incorrectly with respect to either stimulus (fails to press to the positive or presses to the negative), the stimulus is repeated until he behaves correctly. Rewarding correct 'non-responses' to the negative stimulus ensures that every sequence of initial-plus-correction presentations ends with rewarded behaviour, whether the stimulus is positive or negative, and tends to distribute errors to both stimuli more evenly than the usual non-reward method. Such a negative stimulus reward contingency would be difficult to establish in the conventional 'go—no-go' situation. In the Y-maze, it does exist, but only by virtue of the animal's having to alter its physical location by proceeding into a goal-box. The bar-press situation, in addition to permitting negative stimulus reward, can present the sequence of stimuli at a rate not usually possible in conventional procedures.



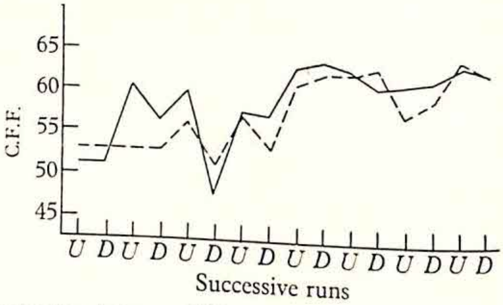


Fig. 7. Psychometric function for a monkey in a discrimination between flickering and steady light. Data obtained with 'symmetrically-rewarded go—no-go method'.

Fig. 8. Successive values of cf.f. for two monkeys. *U*, refers to ascending series; *D*, to descending series. ——, Monkey no. 153; ----, monkey no. 156.

In order to obtain a threshold determination, the value of the positive stimulus is varied either randomly or in a systematic order. Technically, it is simpler to keep the positive stimulus at a fixed value for a single experimental session, varying its value systematically from session to session. A variety of scoring conventions is available; one simple method, the one applied to the data below, is to score the percentage of correct responses to the non-correction (i.e. initial) stimulus presentations. Since at worst, the animal is correct 50% of the time, on the average, the threshold is defined as the 75% correct response point, i.e. the point at which it is making the discrimination half of the time.

This method was developed (in an unpublished study by Mishkin and the author) to determine the critical fusion frequency for normal and brain-operated rhesus monkeys. In fig. 7 is shown a plot of the percentage of correct responses to various values of a flickering light, projected on a diffuse lucite disk just under the bar. Each point represents eighty trials in which the flickering light was randomly alternated with the steady light. The flickering light was generated by a projection lantern and an episcotister with an on/off ratio of 1·00. The 'steady' light was generated by a parallel episcotister constantly presenting a flickering light of 140 c.p.s., with a matched equal brightness of 5·25 foot-candles. The screen diameter was 0·75 in.

While, for this example, the flicker was successively increased until the animal's correct responses approached 50%, in practice it was thought more advantageous not to force the discrimination to 'chance' too frequently, not only for the sake of efficiency, but also to avoid the conditions for classical Pavlovian neurosis. The procedure adopted was to raise the flicker value until the animal received a score of less than 75%, then to lower the value until it received a score of better than 90%. In this way one could come up and down the scale many times without difficulty. In Fig. 8 is shown a plot of the 75% point on successive up and down runs for two rhesus monkeys. Each point is the interpolated value in a series going from above 90% to below 75%, as just described.

The experimental work reported here was conducted at the Psychological Laboratories, Harvard University, and at the Department of Neurophysiology, Institute of Living. Part of the research was supported by Contract DA-49-007-MD-401, Department of the Army, Washington, D.C. The author would like to express his appreciation to Drs Mortimer Mishkin, Karl H. Pribram, and William A. Wilson, Jr., for their helpful suggestions in the preparation of this paper. Mrs Marilyn Tucker and Miss Lila Rupp greatly assisted with technical and mechanical aspects of preparing the manuscript.

REFERENCES

DEUTSCH, J. A. (1951). A preliminary report on a new auditory after-effect. *Quart. J. Exp. Psychol.* **3**, 41–6.

FERSTER, C. B. (1953). The use of the free operant in the analysis of behavior. *Psychol. Bull.* **50**, 263–74.

HEISE, G. A. (1953). Auditory thresholds in the pigeon. *Amer. J. Psychol.* **66**, 1–19.

SKINNER, B. F. (1953). Some contributions of an experimental analysis of behavior to psychology as a whole. *Amer. Psychologist*, **8**, 69–78.

STEVENS, S. S., MORGAN, C. T. & VOLKMAN, J. (1941). Theory of neural quantum in the discrimination of loudness and pitch. *Amer. J. Psychol.* **54**, 315–35.

(*Manuscript received* 14 *May* 1956)

PERCEPTUAL AND PERSONALITY DYNAMICS IN REACTIONS TO AMBIGUITY

By VERNON HAMILTON

*Birkbeck College, University of London**

A battery of tests designed to elicit differential responses to a variety of ambiguous situations was administered to groups of Control and Neurotic subjects. The data show marked individual differences and considerable evidence for intra-individual consistency in avoidance/non-avoidance of ambiguous situations. A high percentage of significant correlations suggests that the tests measure few, closely related variables. Age and intelligence are only minimally involved. The evidence is taken to be consistent with a theory of cognitive control of the environment mediated via perceptual attitudes operative at varying levels of the individual.

To test the influence of controlling and restrictive personality factors on reactions towards ambiguous situations, groups of anxious, hysterical and obsessional psychiatric patients were compared for tendencies to avoid ambiguity. Neurotic subjects were generally shown to avoid ambiguity more than Controls, and Conversion Hysterics and Obsessionals avoided ambiguous situations more than Anxiety States. It is suggested that avoidance of ambiguity serves to avoid anxiety and conflict. The similarity in behaviour between Obsessionals and Conversion Hysterics is interpreted as being due to the functional similarity and aims of the neurotic symptoms of these two clinical groups.

I. INTRODUCTION

'Intolerance of ambiguity' was first conceptualised by Frenkel-Brunswik (1949), who defined it as 'preference for familiarity, symmetry, definiteness and regularity,...a tendency toward black-white solutions, over-simplified dichotomizing, premature, unqualified either/or solutions', etc. These tendencies she considered to be perceptual characteristics found in association with strong attitudes of social and ethnic prejudice, and with repression as a dominant defence mechanism. The high-prejudice group of Frenkel-Brunswik's investigation was also characterized by an inability to think in terms of probability, and by an avoidance of uncertainty 'as accomplished by the narrowing of meaning and the mechanical repetition of set' (1954). This last aspect of the definition would appear to reduce 'intolerance of ambiguity' to a variant of rigidity in the sense in which the latter is used by Luchins (1942) and Rokeach (1948). Eysenck (1954), however, reports over-all low correlations between measures of 'intolerance of ambiguity' and perceptual and conceptual rigidity, so that any obvious similarities between these two variants may be only a function of definition and experimental technique.

'Intolerance of ambiguity' has been variously investigated. Block and Block (1950–51) find the phenomenon expressed in speed of norm establishment in an experiment with the autokinetic phenomenon. Kaplan (1952) reported a contingency between 'intolerance of instability' in the perceptual field induced by means of aniseikonic lenses, and avoidance of equivocality and contradictory material. Further demonstrations of 'intolerance of ambiguity' and its relationship to flexibility of thought (Blum, 1954), need for clarification in ambiguous inter-personal situations (Martin, 1954), and specific antecedent conditions of psychological stress (Smock, 1955*a*, *b*), suggest that the individual's response to ambiguous perceptual situations may depend on a specific organization of personality to deal consistently with perceptual situations lacking structure, definition and certainty.

---

* This investigation forms part of a thesis for the degree of Doctor of Philosophy in Psychology deposited with the University of London Library. It was written while the writer held a Research Studentship at Birkbeck College, University of London, 1953–55.

Experimental evidence in support of a principle of generalized cognitive control has been advanced by Klein and his collaborators (1951*a*, *b*, *c*, 1953*a*, *b*, 1954). According to this theory, cognitive control of the environment is achieved through the operation of 'perceptual attitudes' of levelling, form-boundedness, intolerance of instability and so on.

The aim of the first part of the investigation reported here was twofold. First, to determine if there are consistent individual differences in response to a variety of ambiguous situations. Secondly, to determine whether consistency of response to ambiguity is sufficiently marked to serve as support for the view that consistent perceptual attitudes are instruments of an 'adaptive principle of ego control'.

These aims were formulated in the following questions: (1) Are there individual differences in avoidance of ambiguity as measured by a battery of tests? (2) Are subjects consistent in their responses to ambiguity from test to test? (3) Are differences in avoidance of ambiguity related to differences in age, intelligence, sex or occupation?

(i) *Description of tests*

Eleven test situations were chosen as satisfying the criteria of an ambiguous situation. Only ten of them will be discussed here. The eleventh, apparent movement, will be treated elsewhere in greater detail. Empirical evidence had been obtained in a pilot study suggesting that the tests discriminated sufficiently well between individuals and that significant correlations between them were consistent with the measurement of only a few related traits.

Test 1. *Weight Discrimination.* Sixteen uniform black cans were compared with a can of standard weight. Twelve of the cans were objectively equal in weight to the Standard, two were noticeably lighter and two noticeably heavier. Subjects were told that none of the variables was equal to the Standard, and were required to sort all the weights into the following three categories: 'Heavier', 'Lighter', and 'May be Heavier or Lighter, I can't decide which'.

Test 2. *Length Discrimination.* Sixteen lines on uniform background were compared with a line of standard length. Twelve of the lines were objectively equal to the Standard, two were clearly shorter, and a further two equally clearly longer. Subjects were told that none of the variables were equal to the Standard, and they were required to sort the lines into the following three categories: 'Longer', 'Shorter', 'May be Longer or Shorter, I can't decide which'.

Test 3. *Brightness Discrimination.* Subjects were required to compare sixteen light stimuli with a light of standard brightness. Twelve of the stimuli were of objectively equal brightness to the Standard, two were clearly brighter, and two were clearly dimmer. Subjects were told that none of the stimuli were equal to the Standard, and they were asked to assign their judgements to one of the following three categories: 'Brighter', 'Dimmer', 'May be Brighter or Dimmer, I can't decide which'.

Test 4. *Block Sorting Test.* This consisted of forty-four wooden blocks of varied dimensions and shape. Fourteen were rectangular or triangular, five were circular, and twenty-five incorporated both rectangular and circular features in the same block. The sorting categories in this case were 'Circular', 'Rectangular', 'May be Circular or Rectangular, I can't decide which', and a fourth category, 'To be sorted into separate groups afterwards'.

Test 5. *Ambiguous Drawings I. Sheaf–tree series.* In this, the first of three series of ambiguous drawings, some of the 12 drawings were clearly sheaves, some were clearly trees, while others were indefinite or contained in the same drawing, simultaneously, features characteristic of both trees and sheaves. The sorting categories were: 'Sheaf', 'Tree', and 'May be either a Sheaf or a Tree, I can't decide which'.

Test 6. *Ambiguous drawings II. House–car series.* This was similar to test 5, except that the two objects in the drawings were a house and a car. The sorting categories were: 'House', 'Car', and 'May be either a House or a Car, I can't decide which'.

Test. 7. *Ambiguous Drawings III. Car–dog series.* This series was different from tests 5 and 6 because, in this case, none of the drawings incorporated features of both objects at the same time. The intention in most of the drawings here was to present rather vague or indefinite designs for classification. The categories were: 'Dog', 'Car', and 'May be either a Car or a Dog, I can't decide which'.

Each series was presented in random order, with all the drawings of each series being laid out in front of the subject at the same time. This procedure was adopted to counteract any possibility of an 'Einstellung' effect. On completion of a series, subjects were asked to re-sort the drawings which they had placed into the middle category into an 'Either' group and a group which consisted of drawings which were considered to be really 'Neither' object 1 nor object 2.

Test 8. '*Ambiguity*' *scale.* This scale set out to measure those aspects of an ambiguous situation which will be defined as lack of absoluteness, indefiniteness, and equivocality of interpretation, with special reference to social situations. There were thirty-three items, ten of them 'dummies', which had to be 'posted' into boxes marked 'True' and 'False'. The following are two examples of the type of statement on which judgement had to be made. 'A solution to a personal problem cannot be both right and wrong—it is either one or the other.' 'I hate leaving a decision to hang in the balance—I feel I simply have to decide either one way or the other.'

Test 9. *Reversible Perspectives I. Rubin figure.* Oscillations of the 'Double Cross' figure were measured by an electric counter operated by the subject, over two periods of 30 sec each. The oscillations were measured under two experimental conditions: (1) autonomous reversals, subject passive; and (2) controlled reversals with subject attempting to reverse as often as possible. Under both conditions, subjects were required to fixate the centre point of the design and to control blinking.

Test 10. *Reversible Perspectives II. Necker cube.* Conditions and design were identical with test 9. The calculated reliability of reversals under autonomous conditions for the two reversible figures was 0·844 and 0·876 respectively.

These tests were administered to 102 subjects. Data concerning the constitution of the experimental groups are found in Table 3.

(ii) *Results of correlational study*

Correlation coefficients were calculated using Spearman's Rho formula, corrected for ties by the formula given by Kendall (1948). This was considered to be the most economic method of handling the data, and since no assumptions about the normality of the distributions were made, it may be considered valid. Garrett (1947) and Kendall (1948) have demonstrated that, with a sizeable sample, Rho does not differ materially from Pearson-Bravais r.

Table 1. *Means, standard deviations and range of tests*

	Discrimination tests, C.D. scores			Blocks		Ambiguous drawings, 'Either' scores				Autonomous Reversals		Controlled Reversals	
Test	Weight	Length	B'ness	C.D.	'Others'	Sheaf–tree	House–car	Car–dog	Ambiguity scale	Rubin	Cube	Rubin	Cube
Mean	2·76	3·95	3·84	5·58	19·68	1·13	0·98	1·01	9·77	9·89	9·95	19·32	17·29
S.D.	2·23	2·99	2·72	6·31	11·47	1·38	1·37	1·32	3·43	8·21	5·12	11·93	7·28
Range	0–11	0–12	0–12	0–27	1–43	0–7	0–5	0–5	1–17	0–49	1–26	0–65	0–56

C.D. = 'Can't Decide' scores. 'Others' = blocks which are neither rectangular, nor circular, nor C.D.

Table 1 shows that the data were rarely normally distributed. The results obtained from the Discrimination Tests indicate that individual differences in 'Can't Decide' (C.D.) responses are quite considerable. It appears that in no case did the number of

C.D. responses exceed 12. This suggests that the four objectively unequal stimuli in each test were in fact perceived as being different. Table 2 indicates that a considerable percentage of inter-test correlations is significant at the 1% or 5% level, and that more than 75% of them are above ±0·100. From the matrix a fairly clear picture is obtained of how the tests are grouped together. Two groups may thus be distinguished: the first consisting of the Discrimination Tests, the Block Sorting Test and the series of Ambiguous Drawings; the other of the tests of Reversible Perspective and the 'Ambiguity' Scale. The two groups are interrelated by virtue of the correlations between the Ambiguous Drawings and the Ambiguity Scale, and the correlations between the Discrimination Tests and the tests of Reversible Perspective.

In order to exclude the possibility that the relatively consistent significant correlations between the tests could be accounted for in terms of consistently high correlations for one or two of the experimental subgroups only, separate correlation matrices for each of the experimental subgroups were obtained for seven of the most highly correlated tests. Comparisons of correlation coefficients, obtained from each of the experimental groups separately, indicate that none of the 105 correlations is significantly different from any other. On the other hand, the expected over-all decrease in the size of the correlation coefficients with the increase in homogeneity of the smaller samples materialized in only 68% of the new correlations. This finding suggested that the heterogeneity of the subsamples with respect to the measured behaviour of avoidance of ambiguity may be considerable, and that, on some of the tests at least, the differences between some of the experimental subgroups might well not be significant.

(iii) *Discussion*

Some form of relationship between the variable tapped by the tests, and intelligence and age was bound to be demonstrated, since there are few forms of behaviour, especially in the cognitive sphere, which cannot be shown to vary with both intelligence and age. In fact, it might be thought that intelligence is maximally involved in any process requiring perceptual judgements. In the present study, intelligence correlated with the 'Ambiguity' Scale and the Rubin figure under both autonomous and controlled conditions of oscillation. It might therefore be held that the correlation demonstrated between the Scale and Reversible Perspective is spurious. Partial correlation invalidates this argument ($r_{12\cdot3} = -0\cdot284$ and $-0\cdot293$) and shows that the tests are still significantly related when intelligence is held constant.

The position for age is similar. Evidence has accumulated (Welford, 1951) to suggest that the older section of the population would tend to have the greater difficulty in dealing with new experiences, and that this deterioration is particularly evident in lack of ability to organize new stimuli and select the relevant ones for differential response. For these reasons age might have had considerable influence in the present investigation. Correlations, however, were shown only with the Scale and with number of oscillations of the Rubin figure under controlled conditions. Partial correlation does not materially detract from the original correlation between these two experimental tests, and the new correlation is still significant at better than the 1% level.

The individual differences in avoidance of ambiguity, the consistency of the behaviour in functionally related situations as demonstrated here, are taken to support Klein's

Table 2. *Correlation matrix*

Significance level of r: $N-2$	5%	1%
100	0·195	0·254

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weights, C.D.	1	.	0·425	0·385	0·201	−0·130	0·381	0·325	0·357	0·158	0·204	0·231	0·201	0·204	0·057	0·020
Length, C.D.	2	.	.	0·538	0·219	−0·041	0·353	0·408	·0374	−0·061	0·312	0·122	0·265	0·123	0·130	0·083
Brightness, C.D.	3	.	.	.	0·273	−0·195	0·195	0·216	0·278	0	0·195	0·070	0·221	0·170	0·120	−0·103
Blocks, C.D.	4	.	.	.	.	.	0·296	0·195	0·184	−0·050	0·010	−0·040	0·020	0·010	−0·003	0·030
Blocks, 'Others'	5	.	.	.	.	.	−0·293	−0·256	−0·253	−0·030	0·040	0·130	0·030	0·090	0·020	0·030
Sheaf–tree	6	.	.	.	.	.	.	0·661	0·585	0·285	0·202	0·175	0·120	0·135	0·060	−0·032
House–Car	7	.	.	.	.	.	.	.	0·591	0·304	0·195	0·150	0·146	0·067	0·060	−0·056
Car–Dog	8	.	.	.	.	.	.	.	.	0·232	0·095	0·083	0·044	0·118	−0·020	0·021
'A' Scale*	9	.	.	.	.	.	.	.	.	.	0·340	0·286	0·360	0·350	0·262	−0·219
Rubin, Autonomous	10	.	.	.	.	.	.	.	.	.	.	0·539	0·740	0·470	0·300	−0·094
Cube, Autonomous	11	.	.	.	.	.	.	.	.	.	.	.	0·310	0·550	0·072	−0·103
Rubin, Controlled	12	.	.	.	.	.	.	.	.	.	.	.	.	0·560	0·369	−0·266
Cube, Controlled	13	.	.	.	.	.	.	.	.	.	.	.	.	.	0·141	−0·118
Intelligence	14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Age	15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

(* The signs for this test have been reversed).

(1951*a*, 1954) theory of perceptual attitudes and systems of cognitive control. The degree of consistency from situation to situation is strongly in favour of a relatively permanent and pervasive organization corresponding to some kind of need state or motivational 'set' of the individual.

An explanation of the data in terms of differences in avoidance of ambiguity is preferred to a number of *ad hoc* explanations, mainly because it is comprehensive. The data may be taken to show that several of the dichotomies previously isolated by Klein, such as form-boundedness and intolerance of instability and overlap, tend to be correlated. It is posited that, functionally, the consistent perceptual attitude operative in all the situations had, as its object, the control of ambiguity and uncertainty—the control of indefiniteness, unstructuredness and instability to a level required by a central principle of personality organization or motivational need. This explanation is consistent with a theory (Beier, 1951; Cowan, 1952; Smock, 1955*a*, *b*) that specified perceptual behaviour may be predictably changed by modifying the anxiety level of subjects. More light could thus be thrown on the problem at the explanatory level by investigating the responses to ambiguity of individuals differing in a relatively well-known manner in their affective behaviour, or personality structure. This has been attempted in the second part of the investigation.

II. A COMPARISON OF PSYCHIATRIC AND CONTROL GROUPS

(i) *Theoretical considerations*

The original study on 'intolerance of ambiguity' by Frenkel-Brunswik (1949) suggested that differences in this behaviour may be specifically related to the individual's methods of dealing with conflicts and anxiety in the environment. This suggestion finds support in the study by Rosenberg (1953) on the preference for symmetry of Obsessive-Compulsive Neurotics, and the work of Barron & Welch (1952), who reported a significant correlation between preference for symmetry and Hysteria score on the M.M.P.I. It is further supported by the findings of Balken & Masserman (1940) to the effect that Conversion Hysterics and Obsessive-Compulsive Neurotics show a need for certainty, definition, qualifying and classifying in their responses to objectively scored TAT cards. A pilot study briefly reported elsewhere (Hamilton, 1955) indicated that the tests used in the present study were able to differentiate between Control and Neurotic subjects, and also between the Neurotic subgroups themselves. The results led to the search of the clinical literature for experimental and non-experimental evidence consistent with these findings.

Freud (1920, 1926) postulated that there is no neurosis without anxiety and conflict, while Abrahams (1942) considers neurosis a 'flight' from or a compromise to a conflict (cf. also Symonds, 1946; Maslow & Mittelmann, 1941). Thus the Neurotic may be expected to attempt to avoid conflicts or anxiety-arousing situations. The Conversion Hysteric is considered to display this avoidance tendency prominently in his ability to repress and by his facility for producing somatic conversion symptoms (S. Freud, 1893, 1920, 1926, etc.; Brown, 1940; Hadfield, 1950). The Conversion Hysteric's disinclination to face difficulties, his dependence an his ability to dissociate would all appear to be operative in the service of conflict avoidance, an avoidance which may be considered to have its roots in methods first adopted in infancy to deal with undesirable or unacceptable wishes or tendencies. According to Freud (1926), repression may be found to precede the mechanism of reaction formation in the psychogenesis of Obsessionality, a process in which an unacceptable idea is said to become isolated from its affect. Fenichel (1945) deals at length with the similarity between this affective type of isolation and the Obsessional's system-building and classificatory tendencies. Such a mechanism would predispose the Obsessional to classify ideas, events, and objects into mutually exclusive categories; it would lead to relations being considered as being of the 'either...or...' rather than the 'as well as' kind. By forcing the world into his own system, the Obsessional, therefore, retreats from reality just as the Conversion Hysteric does (Fenichel, 1945). In the Anxiety Neurotic these mechanisms are few or non-existent so that he finds it difficult to avoid anxiety and conflict. Where no alternative is available, the Anxiety Neurotic may respond to additional stress by 'leaving the field'. Suggestive evidence from comparative psychology supports the view that conflict consists of incompatible response tendencies, and that neurotic conflict or 'experimental neurosis' may be considered to follow upon 'mutually exclusive, opposite tendencies of relatively absolutely equal strength' (Bitterman, 1944; Miller, 1944; Sears & Hovland, 1941; Russell, 1950; *et al.*). For the purpose of this study, an ambiguous situation was defined as having double or dubious meaning, as being vague, unstructured or subject to fluctuations, as being of doubtful classification, and engendering subjective uncertainty. The tests of the battery were constructed to meet these criteria and to give rise to mutually exclusive response tendencies. On the basis of the above considerations, it was predicted that Neurotics and Controls, and the Neurotic subgroups themselves, might respond to ambiguity in a way that was congruent with their customary manner of handling difficulties, conflicts and situations of doubtful designation in their environment. The test battery consisted of the tests described in detail in the Introduction.

(ii) *Formulation of problem and method*

It was felt that present-day formulations of psychoanalytic theory were too general to permit the formulation and testing of specific hypotheses, and that a scientifically rigorous procedure was probably best served by preceding it by the accumulation of further empirical data. To this end, the problem was phrased in the form of specific questions, to which the empirical data were hoped to supply fairly unequivocal answers. First, what differences are there, if any, in avoidance of ambiguity between a Neurotic and a Control group? Secondly, what differences are there, if any, in avoidance of ambiguity between the Neurotic subgroups? Thirdly, could any group differences in avoidance of ambiguity be accounted for by differences in age and intelligence?

Table 3. *The experimental groups*

	Controls	Anxiety states	Obsessionals	Conversion hysterics	Neurotics
Men	18	10	12	11	33
Women	22	12	8	9	29
Mean age	30·7	33·4	33·7	32·9	33·3
Range	17–68	21–50	21–54	22–47	21–54
Mean I.Q.	110·3	112·9	110·6	109·6	111·0
Range	100–130	95–125	90–130	85–130	85–130

The study was carried out with 22 Anxiety States, 20 Obsessionals, 20 Conversion Hysterics and 40 normal Controls. The psychiatric population were voluntary patients in a number of mental hospitals who volunteered for the experiments. Their selection was primarily governed by the need for homogeneity with respect to symptoms, age and intelligence. Table 3 indicates that this object was resonably well achieved with respect to age and intelligence. A rigorous attempt was made to safeguard the reliability of psychiatric diagnosis in each case. They were patients on whose diagnosis most clinicians would have agreed, to judge by the symptomatic criteria of classification for the three Neurotic subgroups which five psychiatrists were asked to give. (Full details of subjects' symptomatology and psychiatrists' diagnostic criteria will be found in Hamilton, 1956.) These criteria show agreement considerably above what one might have expected to find according to some recent estimates (Ash, 1949; Eysenck, 1952). The Control group was

Table 4. *Means and variances*

Test	Controls	Anxiety States	Obsessionals	Conversion Hysterics
Discrimination tests				
Weight, C.D.				
$\bar{x}$	3·75	3·36	1·05	0·95
S.D.	2·32	1·93	1·28	1·41
Length, C.D.				
$\bar{x}$	4·57	5·54	2·30	2·60
S.D.	2·80	3·06	2·37	1·85
Brightness, C.D.				
$\bar{x}$	4·38	5·14	2·20	3·00
S.D.	2·45	3·22	2·38	2·26
Block sorting test				
$\bar{x}$	6·28	7·05	3·10	5·05
S.D.	7·33	7·26	4·62	5·46
'Others'				
$\bar{x}$	18·80	18·73	24·90	17·25
S.D.	11·05	13·45	8·95	11·99
Ambiguous Drawings				
Sheaf-tree				
$\bar{x}$	1·92	1·09	0·55	0·70
S.D.	1·35	1·13	1·15	0·99
House–car				
$\bar{x}$	1·75	0·86	0·05	0·50
S.D.	1·62	0·91	0·22	1·24
Car–dog				
$\bar{x}$	1·77	0·43	0·31	0·80
S.D.	1·57	0·66	0·69	1·17
'Ambiguity Scale'				
$\bar{x}$	8·30	11·55	9·20	11·35
S.D.	3·81	3·03	4·49	3·69
Reversible Perspectives				
Autonomous Revers. Rubin Figure				
$\bar{x}$	12·75	9·95	7·45	6·55
S.D.	9·37	5·57	5·63	5·33
Autonomous Revers. Cube				
$\bar{x}$	11·92	9·63	9·40	6·90
S.D.	5·80	4·38	4·82	5·81
Controlled Revers. Rubin figure				
$\bar{x}$	25·80	17·59	15·65	11·95
S.D.	14·80	11·55	11·14	10·66
Controlled Revers. Cube				
$\bar{x}$	22·52	14·09	16·70	10·95
S.D.	11·56	6·35	8·78	8·18

made up of hospital and college employees of roughly similar ages. Intelligence quotients for the Control group were obtained on the basis of occupations using Vernon's (1949) table of norms.

(iii) *Results of comparative study*

Non-parametric methods were used to calculate the significance of differences between the experimental groups. Over-all differences among all groups were calculated by using a one-way variance analysis by ranks (Kruskal & Wallis, 1952), the results of which approximate to the χ^2 distribution. Significance of differences between the whole Neurotics and the Control group, between the Neurotic subgroups and the Control group, and between the Neurotic sub-groups themselves, were calculated by means of the Mann–Whitney U-test (Mann & Whitney, 1947).

Table 4 gives the means and S.D.'s of each subgroup. The following facts emerge. (1) The means of the Neurotic subgroups, especially the means of the Conversion Hysteric and Obsessional groups, are generally lower than the means of the Control group (higher in the case of the 'Ambiguity' scale, where a high score is indicative of avoidance of ambiguity). (2) The means of the groups of Conversion Hysterics and Obsessionals are generally lower than the means of the group of Anxiety States. (3) The means of the groups of Conversion Hysterics and Obsessionals do not appear to differ substantially from one another. (4) The distributions of the Neurotic subgroups on almost all tests are highly skewed, whereas those of the Control group show a greater tendency to either normality or bi-modality.

Table 5

	A. Over-all differences among experimental groups		B. Differences between Neurotic and Control groups ($m = 40$, $n = 62$)	
Test	χ^2	P(D.F. = 3)	U	P
Discrimination Tests				
Weight, C.D.	13·1	<0·01	628½	<0·01
Length, C.D.	18·6	<0·001	936	<0·05
Brightness, C.D.	14·8	<0·01	908	<0·05
Ambiguous Drawings				
Sheaf–tree, 'Either'	10·3	<0·02	819	<0·01
House–car, 'Either'	19·7	<0·001	608½	<0·01
Car–dog, 'Either'	2·3	N.S.	644	<0·01
Block Sorting Test				
C.D.	10·1	<0·02	1062½	N.S.
'Others'	5·5	N.S.	1350½	N.S.
Ambiguity Scale	12·9	<0·01	746	<0·01
Autonomous Reversals				
Rubin Figure	17·8	<0·001	761	<0·01
Necker Cube	23·9	<0·001	737	<0·01
Controlled Reversals				
Rubin Figure	20·2	<0·001	653	<0·01
Necker Cube	22·6	<0·001	629	<0·01

Table 5A gives the results of the one-way analysis of variance by ranks. In only two instances are over-all differences among the four groups not shown to be significant. Subsequent individual tests of significance, however, limit the lack of significance shown here to specific groups (viz. Table 6). Table 5B gives the results of tests of significance between the data from all Neurotic subgroups pooled, and the Control group. In only two cases is significance of difference contra-indicated by the U-test.

Table 6 gives the results of tests of significance by the U-technique, between all the groups individually. These substantially confirm the differences between Controls and Neurotics indicated above. They are, however, more important for the comparisons between the Neurotic subgroups. These show that, of the thirteen comparisons between data from the groups of Conversion Hysterics and Anxiety States, six are significant at the 5 % level or better. The comparisons between Anxiety States and Obsessionals yielded seven significant results, while the comparisons between Conversion Hysterics and Obsessionals yielded only three significant differences, one of which—differences on the fourth category of the Block Sorting Test: 'To be sorted into other groups afterwards'—was specifically expected, in that this category was included because it was thought to differentiate between the classificatory tendencies of Conversion Hysterics and Obsessionals.

Table 6. *Inter-group tests of significance by U-technique*

(Controls (C) = 40, Anxiety States (A) = 22, Obsessionals (O) = 20, Conversion Hysterics (H) = 20.)

	Groups compared					
Tests	C *v.* A	C *v.* O	C *v.* H	A *v.* O	A *v.* H	O *v.* H
Discrimination Tests						
Weights, C.D.	—	121*	134*	66*	73*	—
Length, C.D.	—	204*	219*	77½*	84*	—
Brightness, C.D.	—	194*	255	102½*	135	—
Block Sorting Test						
C.D.	—	—	—	133	—	—
'Others'	—	—	—	—	—	121
'Ambiguity' Scale	262½	—	133½*	—	—	—
Ambiguous Drawings						
Sheaf–tree	—	216*	258	141½	—	—
House–car	305½	129*	174*	107½*	—	—
Car–dog	210*	180*	254*	—	—	—
Autonomous Reversals						
Rubin Figure	—	127*	156½*	142	111½*	—
Necker Cube	299½	—	159*	—	126½	127
Controlled Reversals						
Rubin Figure	—	196½*	145½*	—	127½*	—
Necker Cube	225½*	265	138*	—	—	110½

* Denotes significance at 1 % level or better, plain figures denote significance at 5 % level or better. Non-significant data are omitted.

(iv) *Discussion of results*

Two findings are particularly worth noting in the data of the discrimination tests. These are the significant differences between the groups of Anxiety States and Obsessionals, and the lack of significant differences between Conversion Hysterics and Obsessionals. The Obsessional group produced the lowest means in two of the three discrimination tests, that is to say they produced the smallest number of 'Can't Decide' responses. In view of what has been written about obsessional doubt, it can only be assumed (*a*) that it is not operative in situations like those of the tests used here, when a definite judgement into one of three categories has to be made, or (*b*) that it is manifested only in the time taken to reach a decision. Alternative (*b*) may be rejected because it was shown in the pilot study that there were no significant differences in length of judgement time between four relatively homogeneous clinical groups including Obsessionals. It would appear more likely that obsessional doubt does not prevent the individual from making a definite judgement, even if, as in the present case, a correct judgement is objectively impossible. Furthermore, there would appear to be grounds for linking avoidance of the middle category of these discrimination tests with the level of overt anxiety. In two of the three tests the group of Anxiety States had a higher mean of 'Can't Decide' scores than the Control group. There are no grounds for doubting the accuracy of the introspective remarks by a number of mainly anxious patients to the effect that they preferred to make a C.D. judgement because they were afraid of making a mistake. By contrast, Conversion Hysteric and Obsessional patients reported that they found it relatively easy to reach a decision, and that, in any case, they preferred to decide.

The Block Sorting Test was especially devised not only for the purpose of providing a further ambiguous judgement situation, in which wooden blocks incorporating both rectangular and circular features simultaneously in the same stimulus had to be sorted, but particularly in order to attempt to differentiate between the classifying tendencies of Conversion Hysterics and those of Obsessionals. Clinical evidence (Fenichel, 1945), experimental evidence (Rosenberg, 1953), and common everyday knowledge have typed the Obsessional as one who is prone to differentiate sometimes on minutiae, and the Hysteric as one who tends to gloss over sometimes quite marked discrepancies in his environment (Davis, 1948). In the present context that suggested that the Block Sorting Test with two middle categories, 'May be circular or rectangular, I can't decide which' and 'To be sorted into special groups afterwards', together with the 'Circular' and 'Rectangular' categories, might differentiate between these two clinical groups. This was indeed the case. As before, Obsessionals and Conversion Hysterics both tended to avoid the 'Can't Decide' category, but they were significantly differentiated by their scores on the fourth ('Others') category of the test. If, as it appears from the results of almost all the tests of the present battery, Conversion Hysterics and Obsessionals have specifically marked tendencies to classify their environment, indicated either by their avoidance of the middle categories of a response series, or by avoidance of a response that might leave them with an excessive amount of subjective uncertainty, the results of the Block Sorting Test would seem to supply some explanation of this similarity of behaviour. For it would seem that Conversion Hysterics classify by excessive use of few, mutually exclusive categories, while Obsessionals achieve a similar object by classifying into a large number of small categories. Further support for this suggested explanation comes from the series of Ambiguous Drawings. These series of amorphous or compounded drawings had to be sorted initially into three categories: 'Object 1', 'Object 2', and 'May be either Object 1 or Object 2, I can't decide which'. Subjects were then requested to subdivide the middle category into a group of 'Eithers' and a group of 'Neithers'. The result was that the Obsessional group put into the 'Either' category all those designs which, given opportunity, they would have classified in the first instance as 'Neither' Object 1 nor Object 2. They clearly would have preferred to split up these designs with a 'bad fit' into subsidiary groups, just as they had preferred to do in the case of the Block Sorting Test. The rank correlation between 'Others' in the Block Sorting Test and 'Neithers' in the Ambiguous Drawings was $+0{\cdot}161$, which, while falling short of significance at the 5% level, clearly supports the argument.

It may be felt that these findings underline once again the incompleteness of any general theory of cognition which ignores individual differences in the handling of affect. The need to take these into account appears to be barely recognized by Bruner (1956). Bruner's book, '*A Study of Thinking*', had not appeared at the time of writing. It would seem that one of the objects in thinking strategies in concept formation is to minimize surprise or cognitive strain. But these are considered to be functions of environmental cues rather than of motivational needs of the individual. Moreover, what Bruner had called the availability of categories may be determined fully neither by specific cognitive strategies nor principles of cognitive control (Klein, 1954) nor even intelligence, but by individual differences in willingness to accept an equivocal, undetermined, unstructured and subjectively uncertain situation, which may be engendered when no category for categorizing environmental cues is available.

On the whole, the 'Ambiguity' scale differentiates less well between the clinical sub-groups. While Tables 5A and 5B show differences for this test to be significant, Table 6 indicates that this finding is mainly an artefact of the low scores of the Control group (low scores on the scale being indicative of non-avoidance of ambiguity). Special factors might possibly account for the results from the groups of Anxiety States and Obsessionals. The latter score close to the Controls, the former farthest away from them. The scores of Convention Hysterics and Anxiety States are also hardly differentiated. This lack of selective sensitivity is not entirely maintained when the items carrying a group's highest scores are examined, because several items are then seen to differentiate very well. Furthermore, the groups of Conversion Hysterics and Obsessionals showed a greater amount of inconsistency and contradiction in their responses, which might be endogenous to their personality organization—reflecting what has been called 'the closeness of opposites' (Frenkel-Brunswik, 1951; Rokeach, 1954).

The results of the tests of Reversible Perspective show clear differences between Neurotic and Control subjects as already reported by Eysenck (1947). The Conversion Hysterics of the present study recorded the smallest number of oscillations both under controlled and autonomous conditions of the experiment. The difference between them and the Obsessionals is significant in only two cases out of four, while the difference between Obsessionals and Anxiety States is significant or approaches significance also in two out of four cases (viz. Table 6). On balance, therefore, a more careful and analytic study might be expected to show considerable similarity between the responses of Obsessionals and Conversion Hysterics to the Reversible Perspective situation. This expectation derives reinforcement from the correlation of these tests with the Discrimination Tests and the Ambiguous Drawings, on which a great deal of affinity between these clinical groups was actually demonstrated.

III. CONCLUSIONS

In attempting to account for the findings, some more obvious explanations must first be considered. It has already been shown that age and intelligence are only minimally involved. Though both correlate with some tests that are themselves correlated, these correlations remain significant after partial correlations have been calculated. Table 3 shows that age and intelligence are as closely matched as possible in an investigation whose prime concern was with obtaining clear-cut representatives of specific clinical groups.

A rather obvious explanation may be suggested in terms of differences in suggestibility. While Eysenck (1947) found fairly convincing differences between Neurotics and

Controls, showing the former to be more suggestible than the latter, three objections must be raised against considering suggestibility as an adequate explanation of consistent avoidance of ambiguity.

(1) There is no evidence to show that suggestion as contained in the body-sway type of test and as contained in the instructions of the present battery is equally likely to arouse compliance.

(2) If suggestion as contained within the test instructions was, however, operative, affecting the Neurotic group more than the Controls, then the responses of the Neurotic group to the tests of Reversible Perspective should have yielded higher scores than they in fact did, because in this case suggestions of reversibility of design were quite specific, while they were toned down and controlled in the other tests.

(3) None of the published evidence on differences in suggestibility between Neurotic subgroups is conclusive on the question as to which group is more suggestible. Clinical and medical evidence favours Hysterics (McDougall, 1929), Eysenck's evidence (1947) favours Dysthymics, without distinguishing between Obsessionals and Anxiety States. Neither theory is adequate to explain similarity of performance of Conversion Hysterics *and Obsessionals*.

It may be argued that group differences were due to different levels of aspiration in the subjects. Anxious patients are known to be more highly motivated in terms of goal discrepancy scores (Himmelweit, 1947). In the present series of tests this analogy is inapplicable. Subjects do not know what constitutes a 'good' performance and there is no knowledge of results. On the other hand, it would be surprising indeed if there were no differences in subjects' attitude to the testing situation or even to the investigator. This, however, is a variable that is not easily conducive to measurement, and besides, there are no grounds for supposing that it should be found to operate equally in all tests.

Explanations in terms of differences in perseverative tendencies would also have to be of an *ad hoc* kind, because they could apply only to some and not to other tests, while in some of the situations, viz. Weight Discrimination and Ambiguous Drawings, their possible effects were balanced out. Moreover, Eysenck (1947) found no significant differences in perseveration between Hysterics and Dysthymics. This could account for lack of differences between groups of Conversion Hysterics and Obsessionals, but not for differences found here between Anxiety States and Conversion Hysterics and Obsessionals. In addition, there is a further difficulty in that Rim's study (1955) indicates that a battery of tests of perseveration was not analysable into a general factor of perseveration.

The explanation preferred here is one that is economical, comprehensive, consistent with experimental evidence and in harmony with psychopathology. The clinical groups of this investigation are known to differ with respect to their handling of conflict and anxiety. Conversion Hysterics are known to repress both. Obsessionals are considered to have passed through a phase of repression before obsessive-compulsive symptoms have been developed in which anxiety is being 'bound'. Such mechanisms have been copiously described as defence mechanisms of the ego (S. Freud, 1893, 1894, 1896, 1915, 1936, etc.; A. Freud, 1942; Brown, 1940; Fenichel, 1945) and reference has already been made to them. From their descriptions emerges the picture of the neurotic individual who has a need to avoid 'dangerous' situations. Ego defence mechanisms may be conceived of as operating in the service of this need. Limitations placed upon behaviour such as obsessive-compulsive symptoms, avoidance of difficulties, rigid classification of the environment, scotomization and physical conversion symptoms, which appear in the process of controlling anxiety, would all seem to be designed to reduce the chances of further conflicts and anxiety having to be faced by the subject. It may be considered an attempt to maintain the individual in an environment that is predictable, and in which meaning of stimulus and outcome of response are subjectively certain. This quality, however, is markedly absent from a conflict and anxiety situation. They are qualities, too, which are markedly absent from ambiguous situations. Miller (1951) considers that the development of protection against ignorance, i.e. rapid closure behaviour, and the tendency to achieve environmental stability as demonstrated, for instance, in the 'distorted room phenomena', are factors to lessen fear or anxiety that might arise 'from the constant threat of an uncertain environment'.

It is in this sense that the situations of the present battery of tests may be perceived as threatening by some subjects more than by others, for they not only present conflict situations, but by virtue of their objective lack of definition and structure they may give rise to subjective uncertainty and therefore anxiety.

These conceptualizations, only partly based on experimental data, would encourage no more than a tentative formulation of a general principle which would account in economical terms for the great majority of the experimental findings on avoidance of ambiguity, the theoretical considerations bearing on cognitive control, and the known mechanisms of ego defence. This formulation might run as follows. 'Avoidance of ambiguity as a principle and expression of cognitive control is found in association with a relatively high degree of total anxiety, but particularly where the principal defence mechanism adopted by the individual to cope with anxiety and conflicts is repression. This mechanism leads the individual to deny reality rather than acknowledge it. It becomes generalized to the perceptual field of operation, where by negative methods of limiting and restricting the individual's field of awareness and behaviour, it tends to lead to the avoidance of responses which might result in uncertainty and anxiety, on account of the degree of perceptual conflict, equivocality and unstructuredness inherent in such situations. By avoiding ambiguity, the Neurotic person, and the Conversion Hysteric and Obsessional in particular, would appear to avoid both subjective uncertainty and conflictful situations. By avoiding uncertainty and conflict, the individual would appear to avoid further anxiety.'

The work was greatly facilitated by the generous help and co-operation of the physicians-in-charge and the clinical psychologists of the following Hospitals: The Cassell, Belmont, Banstead, St Ebba's, Horton and West Park. I wish also to acknowledge the assistance extended to me by Mr B. M. Foss, Mr H. M. B. Hurwitz, Dr J. Brown and Miss T. Veness, all of the Psychology Department, Birkbeck College, and of Dr S. G. Lee, of Bedford College, University of London. The investigation was supervised by Prof. C. A. Mace and benefited substantially from his support. Financial aid was provided by the Governing Body of Birkbeck College, and the Central Research Fund of the University of London.

References

Abrahams, K. (1942). *Selected Papers on Psycho-Analysis.* London: Hogarth.

Ash, P. (1949). The reliability of psychiatric diagnosis. *J. Abnorm. Soc. Psychol.* **44**, 272–6.

Balken, E. R. & Masserman, J. H. (1940). The language of phantasy. III. The language of phantasies of patients with conversion hysteria, anxiety state and obsessive-compulsive neurosis *J. Psychol.* **10**, 75–86.

Barron, F. & Welch, G. S. (1952). Artistic perceptions as a possible factor in personality style. *J. Psychol.* **33**, 199–203.

Beier, E. G. (1951). The effects of induced anxiety on flexibility of intellectual functioning. *Psychol. Monogr.* **65**, no. 9.

Bitterman, M. E. (1944). Behaviour disorder as a function of the relative strength of antagonistic response tendencies. *Psychol. Rev.* **51**, 375–8.

Block, J. & Block J. (1950–51). An investigation of the relationship between intolerance of ambiguity and ethnocentrism. *J. Personality,* **19**, 303–11.

Blum, A. (1954). A technique for measuring rigidity-flexibility in children and adults. *Amer. Psychol.* **9**, 334. (Abstract.)

Brown, J. F. (1930). *The Psychodynamics of Abnormal Behaviour.* New York: McGraw-Hill.

Bruner, J. S. (1956). Strategies in thinking. *Bull. Brit. Psychol. Soc.* **28**, 73–4. (Abstract.)

Cowan, E. L. (1952). The influence of varying degrees of psychological stress on problem solving rigidity. *J. Abnorm. Soc. Psychol.* **47**, 512–19.

Davis, R. D. (1948). *Pilot Error.* Air Publications 3139a. London: H.M.S.O.

Eysenck, H. J. (1947). *Dimensions of Personality.* London: Kegan Paul.

Eysenck, H. J. (1952). *The Scientific Study of Personality.* London: Routledge & Kegan Paul.

Eysenck, H. J. (1954). *The Psychology of Politics.* London: Routledge and Kegan Paul.

Fenichel, O. (1945). *The Psycho-Analytic Theory of Neurosis.* New York: Norton.

Frenkel-Brunswik, E. (1949). Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Personality,* **18**, 108–43.

Frenkel-Brunswik, E. (1951). Personality theory and perception. In *Perception: An Approach to Personality.* Ed. Blake, R. R. and Ramsay, G. V. New York: Ronald.

Frenkel-Brunswik, E. (1954). Further explorations by a contributor to 'The authoritarian personality'. In *Studies in the Scope and Method of 'The Authoritarian Personality'.* Continuities in Social Research. Ed. Christie, R. and Jahoda, M. New York: Free Press.

Freud, A. (1942). *The Ego and the Mechanisms of Defence.* London: Hogarth.

Freud, S. (1893). On the psychical mechanism of hysterical phenomena. (With Breuer). In *Collected Papers,* I. London: Hogarth.

Freud, S. (1894) The defence neuro-psychosis. In *Collected Papers,* I. London: Hogarth, 1940.

Freud, S. (1896). The aetiology of hysteria. In *Collected Papers,* IV. London: Hogarth, 1940.

Freud, S. (1915). Repression. In *Collected Papers,* IV. London: Hogarth, 1940.

Freud, S. (1920). *A General Introduction to Psycho-Analysis.* New York: Boni and Livertight.

Freud, S. (1936). Inhibitions, symptoms and anxiety. London: Hogarth.

Garrett, H. E. (1947). *Statistics in Psychology and Education.* New York: Longmans, Green.

Hadfield, J. A. (1950). *Psychology and Mental Health.* London: Allan and Unwin.

Hamilton, V. (1955). Reactions to ambiguity as a personality variable. *Bull. Brit. Psychol. Soc.* **26**, 18. (Abstract.)

Hamilton, V. J. (1956). Reaction to uncertainty and ambiguity as a personality variable. Ph.D. Thesis. University of London Library.

Himmelweit, H. T. (1947). The comparative study of the level of aspiration of normal and neurotic persons. *Brit. J. Psychol.* **37**, 41–59.

Kaplan, M. (1952). Predicting memory behaviour from cognitive attitudes towards instability. *Amer. Psychol,* **7**, 322. (Abstract.)

Kendall, M. G. (1948). *Rank Correlation Methods.* London: Griffin.

Klein, G. S. (1951*a*). The personal world through perception. In *Perception: An Approach to Personality.* Ed. Blake, R. R. and Ramsay, G. V. New York: Ronald.

Klein, G. S., Schlesinger, H. & Meister, D. E. (1951*b*). The effect of personal values on perception: an experimental critique. *Psychol. Rev.,* **58**, 96–112.

KLEIN, G. S. & SCHLESINGER, H. (1951*c*). Perceptual attitudes towards instability: I. Prediction of apparent movement experiences from Rorschach responses. *J. Personality*, **19**, 289–302.

KLEIN, G. S. (1953*a*). *The Menninger Foundation Research on Perception and Personality*: a review. *Bull. Menninger Clinic*, **17**, no. 3.

KLEIN, G. S. & SMITH, G. J. W. (1953*b*). Cognitive controls in serial behaviour patterns. *J. Personality*, **22**, 188–213.

KLEIN, G. S. (1954). Need and regulation. In *Nebraska Symposium on Motivation*. Ed. Jones, M. R. Lincoln: University of Nebraska Press.

KRUSKAL, W. M. & WALLIS, W. A. (1952). Use of ranks in one-criterion variance analysis. *J. Amer. Statist. Ass.* **47**, 583–621.

LUCHINS, A. S. (1942). Mechanisation in problem solving. *Psychol. Monogr.* **54**, no. 6.

MANN, H. B. & WHITNEY, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* **18**, 50–60.

MARTIN, B. (1953–54). Intolerance of ambiguity in inter-personal and perceptual behaviour. *J. Personality*, **22**, 494–503.

MASLOW, A. H. & MITTELMANN, B. (1941). *Principles of Abnormal Psychology*. New York: Harper.

McDOUGALL, W. (1929). Suggestion. In *Encyclopaedia Britannica*, XIVth ed., pp. 201–7.

MILLER, J. G. (1951). Unconscious processes and perception. In *Perception: An Approach to Personality*. Ed. Blake, R. R. and Ramsay, G. V. New York: Ronald.

MILLER, N. E. (1944). Experimental studies of conflict In *Personality and the Behaviour Disorders*. Ed. Hunt, J. McV. New York: Ronald.

RIM, Y. (1955). Perseveration and fluency as measures of introversion-extraversion in abnormal subjects. *J. Personality*, **23**, 325–34.

ROKEACH, M. (1948). Generalized mental rigidity as a factor in ethno-centrism. *J. Abnorm. Soc. Psychol.* **43**, 259–78.

ROKEACH, M. (1954). The nature and meaning of dogmatism. *Psychol. Rev.* **61**, 194–204.

ROSENBERG, B. G. (1953). Compulsiveness as a determinant in selected cognitive-perceptual performances. *J. Personality*, **21**, 506–16.

RUSSELL, R. (1950). The comparative study of 'Conflict' or 'experimental neurosis'. *Brit. J. Psychol.* **41**, 95–108.

SEARS, R. R., & HOVLAND, C. I. (1941). Experiments on motor conflict: II. Determination of mode or resolution by comparative strengths of conflicting responses *J. Exp. Psychol.* **28**, 280–6.

SMOCK, C. D. (1955*a*). The influence of psychological stress on the 'intolerance of ambiguity'. *J. Abnorm. Soc. Psychol.* **50**, 177–82.

SMOCK, C. D. (1955*b*). The influence of stress on the perception of incongruity. *J. Abnorm. Soc. Psychol.* **50**, 354–6.

SYMONDS, P. M. (1946). *The Dynamics of Human Adjustment*. New York: Appleton-Century.

VERNON, P. E. (1949). Occupational norms for the 20-Minute progressive matrices test. *Occ. Psychol.* **23**, 58–9.

WELFORD, A. T. *et al.* (1951). *Skill and Age*. Oxford University Press.

(*Manuscript received* 18 *April* 1956)

FREQUENCY, RECENCY, ASSOCIATION VALUE AND TACHISTOSCOPIC IDENTIFICATION

By G. W. H. LEYTHAM

Department of Psychology, University of Liverpool

Studies of the relationship between cognitive factors and the tachistoscopic identification of verbal material (arising out of the failure to control the familiarity of words used in early experiments on personality and perception) have shown that frequency, both relative (Howes & Solomon, 1951) and controlled (Solomon & Postman, 1952), is inversely related to 'recognition thresholds'. Vanderplas (1953), following gestalt principles, found that when frequency was held constant organized material was easier to recognize later than non-organized. Other theories have also been applied in this field, e.g. information theory (Miller, Bruner & Postman, 1954), and learning theory (Reece, 1954). There still seems to be room for more exploratory studies, however, especially of the recency factor, and of the interrelationships between the various factors. This report describes such a study.

Sixteen students (eight men and eight women) acted as subjects. The material consisted of two groups of nonsense syllables, one drawn from Glaze's (1928) 100% association value list (High) and the other from his 20% list (Low). The syllables, in the order in which they were learned, were as follows:

High: rom, tak, bur, giv, dep, soc, haz, puf, liq, ves.

Low: qus, kig, saj, fep, gok, ruw, jih, xaf, leb, zor.

The factorial design of the experiment involved the variables of frequency (One and Four presentations), recency (Immediate and Delayed recognition), type of material (High and Low value syllables), and sex differences. One subject was randomly assigned to each of the resulting sixteen conditions. As only one list of syllables was learned, but both were later identified, each subject acted as his own control. Replication was in terms of the material.

The subjects, who had no prior acquaintance with nonsense syllables, came individually to each part of the experiment. In the learning period, they were asked to memorize the syllables, which were presented in serial order by means of an Elite projector. Each syllable was exposed for 4 sec. and subjects were told how often they would see them. When this was four times, there was an interval of 32 sec. between presentations. The tachistoscopic procedure, which was the same for all subjects, took place 5 min. or 1 week later. The 5 min. period was occupied by reading. A projection tachistoscope was used. Three practice syllables (col, mic, wef) were presented first to overcome the learning effect of the procedure, and to establish individual shutter speed and initial illumination. Each of the original twenty syllables, in a pre-arranged random order, was then presented by the ascending method of limits until correctly reported. Exposure speed was constant, and illumination was increased by 2 volts each trial. The lowest illumination at which a subject correctly reported any of the syllables was taken as his first trial, and the score for each syllable was the number of trials required to report it correctly.

No relationship was found between the serial order in which syllables were learned and the scores obtained on them.

Analysis of variance showed that the between-groups' variance estimate was significantly larger than that for within-groups ($F=4\cdot65$, D.F. $=31$ and 288, $P<0\cdot001$), hence each condition was analysed separately. The mean score for both learned and unlearned groups of syllables was calculated for each subject and the difference tested for significance. No significant differences were found for *single* frequency learning, but all were in the expected direction except for two conditions: Low–Immediate–Male, and Low–Delayed–Female. For material seen *four* times, significant and nearly significant differences were as follows: High–Immediate–Male ($t=3\cdot2$, $P<0\cdot01$), High–Immediate–Female ($t=2\cdot35$, $P=0\cdot035$); High–Delayed–Male ($t=2\cdot72$, $P=0\cdot015$); Low–Delayed–Male ($t=3\cdot93$, $P<0\cdot01$), Low–Delayed–Female ($t=2\cdot84$, $P=0\cdot01$).

The last two results were the most noteworthy. Why should the influence of learning low value syllables appear only after a week and not a few minutes later? To test the reliability of this finding, these two conditions were repeated using eight subjects (four men and four women) in each of the two groups. In addition to the tachistoscopic procedure, these subjects were presented with the twenty syllables (in a new random order) by means of a drop-shutter exposure apparatus wired to a voice key and chronotron. Free associations were asked for, and reaction times noted. Allowing 10 sec. for failures to associate within this interval, the range of mean reaction times for each syllable in the high value group was 1·25–2·33 sec., and for those in the low group it was 2·71–5·64 sec. The corresponding medians were 1·15–1·80 and 1·65–4·30 sec. These figures seem to justify the classification of the syllables as far as these subjects are concerned. Their scores on the tachistoscopic presentation also justify the earlier results for these two conditions: for the Immediate group, the mean score for the learned (Low) syllables (3·45) was not significantly different from that for the unlearned (High) ones (3·50). For the Delayed group, the respective means were 3·45 and 4·81, and the difference was again significant ($t=2\cdot77$, $P<0\cdot01$). A comparison between the scores for the two groups shows that the means for learned material are identical, but that the mean for unlearned material is appreciably higher for the Delayed group than it is for the Immediate. This difference is, in fact, significant ($t=2\cdot95$, $P<0\cdot01$).

One way to explain these results is to suppose that the learned syllables had become organized into a conceptual system, and that the set to perceive nonsense syllables (subjects were merely told that nonsense syllables would be flashed at them) had aroused this system and served to retard identification of non-members. It was found in a study by Postman & Leytham (1951), for instance, that when a series of tachistoscopically presented trait names was unexpectedly followed by a noun (matched for familiarity), the number of trials required to identify the noun was greater than it was for the first, non-trait, practice word. The reinforced trait set seems to have delayed the identification of the noun. If, as suggested, the same process operated in the present experiment, why was it not apparent when the test was carried out immediately after learning? One possible explanation is that it takes time for material of low association value to become organized into an effective conceptual system: an incubation period, rather like that associated with creative thinking, is required.

A consideration of how subjects arrived at correct responses in this study, suggests that, for nonsense syllables at least, there are roughly three stages in this process:

(1) seeing the stimulus (sensory threshold), (2) recognizing the stimulus (perceptual or recognition threshold), and (3) remembering what has been perceived long enough to report or record it. Inability to report a stimulus correctly may thus be due to factors operating at any one or more of these stages, and if this process is general for verbal material, such terms as 'perceptual defense' and 'subception' may beg the question. It is for this reason also that the term 'tachistoscopic identification' (Miller *et al.*, 1954) is to be preferred as a description of a subject's correct verbal response in this kind of experiment.

The writer is indebted to Mr J. E. Fisher for preparing the slides and other material used in this research.

REFERENCES

GLAZE, J.A. (1928). The association value of nonsense syllables. *J. Genet. Psychol.* **35**, 255–67.

HOWES, D. H. & SOLOMON, R. L. (1951). Visual duration threshold as a function of word-probability. *J. Exp. Psychol.* **41**, 401–10.

MILLER, G. A. BRUNER, J. S. & POSTMAN, L. (1954). Familiarity of letter sequences and tachistoscopic identification. *J. Gen. Psychol.* **50**, 129–39.

POSTMAN, L. & LEYTHAM, G. W. H. (1951). Perceptual selectivity and ambivalence of stimuli. *J. Personality*, **19**, 390–405.

REECE, M. M. (1954). The effect of shock on recognition thresholds. *J. Abnorm. Soc. Psychol.* **49**, 165–72.

SOLOMON, R. L. & POSTMAN, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *J. Exp. Psychol.* **43**, 195–201.

VANDERPLAS, J. M. (1953). Frequency of experience versus organisation as determinants of visual thresholds. *Amer. J. Psychol.* **66**, 574–83.

(*Manuscript received* 8 *June* 1956)

THE PREDICTION OF EXAMINATION MARKS OF MECHANICAL ENGINEERING STUDENTS AT KING'S COLLEGE, NEWCASTLE

By R. F. GARSIDE

Department of Psychological Medicine, University of Durham

The correlation coefficients between the examination marks of 108 University engineering students and various predictors are given. It is concluded that School Certificate, G.C.E. (Advanced Level) or Higher National Certificate (A1) results provide the best prediction of engineering marks.

I. Introduction

This paper is concerned with the prediction of the examination marks of mechanical engineering students at King's College, Newcastle. Hohne (1955) and Sanders (1948) both concluded that Australian school-leaving examination results gave the best prediction of first-year University engineering marks. A useful summary of the literature regarding University selection in general is given by Dale (1954).

II. Method

The sample consisted of 108 students who had entered King's College in the sessions 1952–3 to 1954–5 inclusive. Some of these students had taken the first-year examination, whereas some had taken the second-year and the remainder had completed their final examinations. The correlations between the three examinations were calculated, and it was found that the second and final examinations correlated highly (0·86), whereas the average correlation between the first and subsequent examinations was only 0·38 (Fisher's *z* method). It was therefore decided to adopt the second or the sum of the second and final examination marks as one criterion and the first examination marks as a separate criterion. Validities of the predictors against these two criteria were then calculated.

The predictors consisted of the ratings of four interviewers, the results of N.I.I.P. tests 90 A (verbal) and 100 A (verbal and non-verbal) and of Peel's test V.S. 23 (non-verbal), School Certificate examination results, average G.C.E. (Advanced Level) marks or average A1 marks of the Higher National Certificate examination. Only first attempts at the three examinations were considered. The Advanced Level subjects included pure and applied mathematics, physics and chemistry only.

The School Certificate results were scored by awarding 3 marks for a distinction (or a very good), 2 marks for a credit and 1 for a pass. All subjects were taken into account. Two additional marks were awarded for every year by which the candidate was younger than 15 years when he took the examination. Two marks were deducted for every year the candidate exceeded 16 years when taking the examination.

Three of the interviewers (A, B and C in Table 1) interviewed the candidates as a panel and had School Certificate, Advanced Level and A1 results to refer to. The remaining interviewer (D), however, relied entirely upon the data obtained at the interview which he conducted by himself.

III. Results

The product moment correlations between the predictors and the criteria are given in Table 1. Except where shown, $N = 33$ for the first examination validities and $N = 94$ for the second and final examination validities.

Both the average validities of the School Certificate, Advanced Level and A1 results are greater than any other validity, but not significantly so in every case. When, however, these two average validities are themselves averaged, then the resulting average validity (0·54) is significantly greater than the average validity of any other predictor.

Table 1

Predictor	Validities			
	Against 1st examination		Against 2nd and final examination	
Interviewer A	0·370*		0·335**	
Interviewer B	0·242		0·318**	
Interviewer C	0·046		0·355**	
Interviewer D	0·379*		0·179	
Test 90 A	0·361*		0·071	
Test 100 A	0·315		−0·006	
Test V.S. 23	0·103		0·052	
School Certificate	0·550** ($N=22$)	Average 0·578**	0·378 ($N=26$)	Average 0·532**
Advanced Level. G.C.E.	0·727 ($N=6$)		0·688** ($N=17$)	
A1, H.N.C.	—		0·561** ($N=27$)	

** denotes a highly significant correlation ($P<0{\cdot}01$); * a significant correlation ($P<0{\cdot}05$). The differences between the School Certificate, Advanced Level and A1 validities are not significant.

IV. Discussion

The results obtained indicate that School Certificate, Advanced Level or A1 results provide the best prediction of the engineering marks in question. This conclusion is in agreement with those of Hohne (1955) and Sanders (1948); their correlation coefficients between first-year engineering marks and the Australian school-leaving examination were about 0·7, whereas their cognitive test coefficients were about 0·5. It is likely that the second and final examinations are reached by only those students who are intellectually capable. This may explain, at least in part, the vanishing second and final examination validities of the three tests. The low interview validities suggest that the interview is of doubtful value as a predictor of University engineering marks. This conclusion is in accordance with previous findings regarding University selection in general.

References

Dale, R. R. (1954). *From School to University.* London: Routledge and Kegan Paul.

Hohne, H. H. (1955). *Success and Failure in Scientific Faculties of the University of Melbourne.* Melbourne: Australian Council for Educational Research.

Sanders, C. (1948). *Student Selection and Academic Success in Australian Universities.* Sydney: Commonwealth Office of Education.

(*Manuscript received* 25 *June* 1956)

SHAPE RECOGNITION IN RATS

By P. C. DODWELL

Department of Psychology, Birkbeck College, University of London

Reasons are given for doubting whether certain of Lashley's experiments, purporting to be investigations of shape discriminations, really are such. In particular his finding concerning the square-circle discrimination cannot be held to lend support to Deutsch's theory of shape recognition. An experiment is reported which shows that rats can discriminate square from circle under certain conditions of pre-training, and do so under these conditions by responding to the horizontal base of the square; the response is transferred to other shapes having a horizontal base, but not to shapes lacking it. There is some evidence to show that this happens for small shapes—where Deutsch's system could be operative—and this is therefore evidence against his theory. A theory of shape recognition is proposed of the same logical type as Deutsch's, which accounts for the results here reported, accords well with other known characteristics of shape recognition in primitive visual systems, and presents less difficulties than Deutsch's system as a possible neurological model.

I. Introduction

A well-known paper by Lashley (1938) sets out some of the basic characteristics of vision in rats. The purpose of the experiments he reports was to establish norms of behaviour, rather than to prove any hypothesis or support any particular theory of vision. One of Lashley's findings, apparently quite general in the experimental situation used, was that the rat tends to isolate from a complex visual stimulus an element, or elements, to serve as the basis of discrimination. The animal uses only part of the information available. This was noted in shape discrimination; under conditions of equal area and brightness, shapes tended to be discriminated from one another by isolation of an element or property present in only one of the shapes. It was also found that the most important cue in this type of discrimination is the relation of the part-figure to the edge of the stimulus card: 'Systematic tests have shown that the majority of animals do not react to the entire figure presented in the stimulus patterns.... The most important factor in defining the part-figure is the relation of the stimulus card to the surrounding frame.' (p. 152.)

Further exposition will be simplified by the definition of three terms: (i) 'stimulus shape' is the pattern drawn on a stimulus card; (ii) 'stimulus card' is the card presented in the jumping stand (Lashley, 1938 p. 126); (iii) 'discriminated shape' is the part of the stimulus situation to which the rat is responding. For (i) and (ii) simply 'stimulus' and 'card' will be used when the context makes the meaning clear.

There are several important features of Lashley's experiments which make their interpretation difficult, especially so far as shape recognition is concerned. (*a*) The rats were pre-trained to discriminate between horizontal and vertical striations. (*b*) The ratio of stimulus shape to stimulus card was large; that is to say, the internal dimensions of the shapes were of the same order of magnitude as the distances between shape and card edges. (*c*) The stimulus shapes presented quite a large visual angle at the rat's eye. The peculiarities mentioned above, concerning part-figure discrimination and discrimination by the relation of stimulus shape to card edge (the latter I shall call a 'boundary effect') may be artefacts of the conditions of the experiment, rather than characteristics of shape discrimination.

It is by no means certain, then, that Lashley was investigating shape discrimination as such; especially as many of the figures used were solid figures, rather than outline figures, it appears possible that the rats were responding to the distribution of light on the stimulus card, and not to shape. (This point will be further elaborated below, in §III.) Even if they were reacting to shape, the discriminated shape was not in general the stimulus shape, or even a part of it, but some relationship between the stimulus shape and the stimulus card.

In a recent paper (Deutsch, 1955), a theory of shape recognition was put forward by Deutsch, which accounts for Lashley's rather surprising finding that rats have great difficulty in discriminating a square from a circle of equal area and brightness. This theory, as it stands, requires that a shape, to be recognized, must all be seen at the same time, or, to put it more precisely, each part of the discriminated contour must excite a separate receptor, or group of receptors, simultaneously. (The division of the contour into separate parts presents no difficulty, so long as the number of parts is small compared with the number of available receptors. In mammalian eyes, at least, this condition would appear to be fulfilled.) It seems plausible to suppose that the size of a discriminated shape will be restricted by the size of the rat's field of distinct vision; this field, according to Lashley, is small, although its precise extent is not known.

It appears, then, that Deutsch's system could not operate in the conditions of Lashley's experiments, and the anomaly concerning square-circle discrimination which it accounts for so neatly may, in this case, have a quite different explanation. In view of the objections listed above it would be desirable to know whether the supposed inability of the rat to discriminate square from circle holds under conditions in which Deutsch's system of shape recognition could function. This can be done in two stages; the first is to eliminate, as far as possible, the boundary effects, and to find out whether (*a*) the inability to discriminate square from circle still obtains for stimulus shapes of the size used by Lashley under conditions of pre-training similar to his, and (*b*) how this discrimination is affected by reducing the size of the stimulus shapes. The second stage consists in an investigation of the discrimination of small shapes without pre-training in any form of shape discrimination.

The experiment reported below deals only with the first stage, but it is hoped that in due course experiments on the second stage, at present in progress, will be reported.

II. THE EXPERIMENT

The purpose of the experiment was to investigate the discrimination of square from circle in the rat, under conditions similar to Lashley's, but avoiding as far as possible the introduction of boundary effects; to investigate the effect of different sizes of stimulus shape on this discrimination; and to determine to what extent pre-training on horizontal and vertical striations affects subsequent discrimination.

A modified form of Lashley's two-card jumping stand was used; the important modifications were: (i) A Y-shaped jumping platform, with arms set at 45° to each other; (ii) stimulus cards placed at 135° to each other, each card thus facing one of the arms of the platform at right angles; (iii) large black stimulus cards, 11 in. square, each card containing a 4 in. door in its centre, fitting flush with the card; (iv) stimulus shapes, cut out in black cardboard, could be attached to either stimulus card by a simple device; (v) each stimulus shape was outlined in white ink. The advantages which these modifications give will be obvious: it is not claimed that the arrangement described in (iii) which is intended to abolish, or reduce, boundary effects makes the shape completely isolated from the edges of the card, or that the inside edges of the card surrounding the door are invisible. It is an improvement, however, on Lashley's original apparatus, so far as investigating shape discrimination is concerned. The stimulus shapes were 'outline' rather than 'solid' figures, as with the latter discrimination might be based simply on the distribution of brightness in the stimulus shape (see §III below). In Deutsch's theory, which deals only with contours, the two types of figure are equivalent.

The shapes used were squares, circles, triangles and pentagons. There were five sizes for each shape, the squares ranged from side $1\frac{1}{2}$ in. (size 1) to side $3\frac{1}{2}$ in. (size 5), by $\frac{1}{2}$ in. steps, and the sizes of the other shapes were determined by equating them for area, in each case, with one of the squares. Thus there was a series of five sizes for every shape.

During training shapes of equal area were always used: however, there are three cues which might still be used as a basis for discrimination, if all the outlines are of equal width. They are: (i) circumference length (which differs for different shapes of equal area); (ii) total brightness; and (iii) what may be called 'subjective' size.

To eliminate (ii) several copies of each size of the shapes were made, each having a different width of outline, and hence differing in total brightness. These were interchanged during training, from time to time, so that total brightness would not serve to distinguish the shapes; the difference in brightness between the shapes was in any case small. Control of (i) and (iii) was effected by showing different *sizes* of the training shapes together after discrimination had been established; in no case had discrimination been based on either of these cues.

The subjects were ten hooded male rats from the colony kept at the Institute of Experimental Psychology at Oxford. (In fact twelve rats were used, as originally reported (Dodwell, 1956), but the results for the other two, although they support the general conclusions here given are not reported as the experimental conditions for these two animals differed in certain respects from the conditions for the rest of the experiment.) At the start of the experiment they were approximately 100 days old.

All the animals were put on a 24 hr feeding schedule (being fed for 2 hr. in 24) a week before the experiment began, and were given opportunity to become familiar with the apparatus. They were then taught to discriminate a wooden door in a black card from a plain black card, and this situation was continued until jumping was well established.

Next, discrimination of horizontal from vertical striations was learned. In this, as in the subsequent discriminations, the criterion of success was twenty consecutive errorless jumps; the shapes were changed randomly, and the rats not allowed to correct errors (thus a wrong jump left would not necessarily mean that at the next jump right would be correct). Each animal had ten jumps per day. Each correct jump is recorded as a trial, each incorrect jump as an error.

On the basis of their performance in this discrimination the rats were divided into three groups (of 4, 4 and 2) roughly equated for number of trials, and errors made.

The training given to each group, together with results, are shown in Table 1.

Table 1

Group	Discrimination		Size of shape	Trials†	Errors†	Criterion reached	Transfers to other sizes
	Positive	Negative					
I*a**	Square	Circle	5	20	0	Yes	4, 3
I*b**	Square	Circle	1	154	56	Yes	2, 3, 4, 5
II	Diamond	Circle	1, 5	164	143	No	
III*	Circle	Square	5	150	165	No	—

* Two rats only.
† Group averages.

Groups I and II were divided, two rats in each being trained on large shapes and two on small. Group I*a* jumped without further training to the large square, group I*b* required considerable training before reaching the criterion, but then transferred to shapes of all the larger sizes. There were no differences within group II.

Circumstances beyond the experimenter's control made it impossible to continue the experiment with groups II and III for as long as would have been desirable, but the results quoted make it clear that, under the given conditions, the discrimination of square from circle is possible, and is dependent on pre-training on horizontal and vertical striations. It would appear probable that the differences between group I on the one hand, and groups II and III on the other, can be explained in terms of the horizontal base-line of the square, since this is an obvious characteristic of the positive pre-training shape which is present in the positive shape for group I but not for groups II and III (group III showed an initial preference for the square also): having learned the horizontal/vertical discrimination the rats transfer to another shape with a horizontal base line. This is confirmed by a series of transfer tests given to group I after the criterion on the square/circle discrimination has been reached.

Table 2 gives the transfer shapes and results; in the third and fourth columns are the numbers of correct jumps for individual animals. The sizes of shapes shown were the sizes on which the animals learned, and triangles and pentagons were shown with one side (base) horizontal. Each test consisted of a series of ten jumps followed by ten jumps on the the square/circle discrimination.

Table 2

Positive	Negative	I*a*	I*b*
(i) Triangle	Circle	10, 10	10, 10
(ii) Pentagon	Circle	10, 10	10, 10
(iii) Square	Triangle	5, 5 (position habit)	9,—
(iv) Square	Pentagon	5, 5 (position habit)	10, 10
(v) Square (large)	Square (small)	7, 8	6, —
(vi) Square	Diamond	8, 7	9, 10

There were no results for one animal on (iii) and (iv) since discrimination broke down before the end of the series. With this one exception the square/circle discrimination was maintained throughout the series. During transfer tests both doors were left unlocked, and the shapes changed in a random sequence. It might be supposed that the remarkable transfer in (i) and (ii) was due to simple avoidance of the circle; this was not so, as a circle shown with a diamond, an inverted triangle or pentagon gave chance scores, together with the typical conflict symptoms of micturition, squealing and long delays in jumping. Transfers (iii) and (iv) are the only ones which differ markedly for the two groups: apparently group I*a* treat any shapes with horizontal base-lines as equivalent (on transfers (iii) and (iv) both animals jumped to the 'preferred side' irrespective of the shapes shown, but both returned to correct discrimination when square and circle were shown together), but there is a suggestion that group I*b* differentiate square from triangle and pentagon with base horizontal.

The most striking conclusion to be drawn from this experiment is that discrimination has evidently been achieved by the isolation of an element of one of the shapes; there can be no doubt that this common element is the horizontal base line. The isolation of this element occurs during the pre-training on horizontal and vertical striations, and explains the initial preference of groups I*a* and III for the square.

The following detailed conclusions can be drawn, some of them only tentatively, since the number of animals used was small: (1) Pre-training on shape discrimination has a very marked effect on subsequent shape discrimination. (2) Under the given conditions square and circle can readily be distinguished, by the isolation of an element in one of the shapes; the rate at which this discrimination is learned depends on the size of the shapes, and possibly also on the size of the pre-training shapes. (3) There is transfer to a variety of shapes which have the isolated element in common, and this effect persists—at least to some extent—for small shapes. (4) There is no confusion of square with diamond, or of either of these with circles.

These results clearly have important implications for Deutsch's theory of shape recognition.

III. DISCUSSION

Deutsch (1955) lists six main facts which a theory of shape recognition must account for. Briefly, these are that shape recognition can be independent of: (*a*) localization or fixation on the centre of the figure; (*b*) the angle of inclination of the figure in the visual field (rotation on sagittal axis); and (*c*) sizes. Furthermore: (*d*) mirror images are confused; (*e*) visually primitive organisms find it hard, if not impossible, to distinguish square from circle; and (*f*) shape discrimination survives the destruction of large parts of the striate cortex. For a discussion of these six facts, the reader is referred to Deutsch's paper (Deutsch, 1955, pp. 30–1). The system proposed by Deutsch does satisfy these six characteristics, but whether they are all characteristics of shape discrimination in visually primitive organisms, such as the rat, is open to doubt. In particular, facts (*b*) and (*e*) are questionable.

Several investigators have found that a slight rotation on the saggittal axis of a two-dimensional shape interferes with its recognition (Fields, 1935; Gellerman, 1933). Fields puts the threshold for this type of disturbance at between 5° and 10° of rotation for the rat. In this respect, Deutsch's system represents a higher capacity for visual abstraction than that of the rat, or even the monkey (Gellerman).

Fact (*e*) is stated as a characteristic of the visual system of the rat on the basis of Lashley's experiments (Lashley, 1938). It has been pointed out above that there are several serious objections to taking Lashley's experiments as investigations of shape discrimination. Even if they were genuine shape discrimination experiments, the stimulus shapes used were too large to be 'seen whole' and therefore his results are irrelevant to Deutsch's theory; this applies to the finding concerning lack of discrimination of square from circle. (It should be mentioned that failure to discriminate square from circle has been observed in the octopus (Boycott & Young, 1956) but these authors have their own explanation for this).

The results of the experiment described above show that square and circle can be distinguished from one another. There are two important differences between the conditions of this experiment and Lashley's: boundary effects have been reduced, if not eliminated, and the shapes used are 'outline' rather than 'solid'.

That *reduction* of some cues (boundary effects) should facilitate a discrimination may seem puzzling. The use of 'outline' rather than 'solid' figures also involves a reduction of cues, and these are cues of a type which might well interfere with shape discrimination as such; it seems likely that the explanation of the difference between Lashley's findings and mine lies here. It is well known that brightness is a 'dominant' cue for the rat, and it could well be that the use of 'solid' shapes predisposes the rat to seek, if not differences of brightness between figures, different distributions of brightness within figures. Thus a rat might distinguish between a (solid) square and a triangle by distinguishing 'bright above and below' from 'bright below'. It would not be possible to discriminate a square from a circle on this basis using above-below or right-left comparison's or comparisons of one or more quadrants in one figure with the same quadrants in the other. It seems plausible to suppose that attempts at discrimination of this sort would be less likely to occur when 'outline' figures are used against a dark, homogeneous, internal and external background. If Lashley's rats were attempting to discriminate on the basis of brightness distribution, their failure is understandable.

However, it remains to be seen whether rats fail to make the square-circle discrimination under conditions in which Deutsch's system could operate. Data for group I*b* reported above suggest that they can make it, but the only really satisfactory test would be in a situation where the rats had had no pre-training on shape discrimination of any sort.

Whether or not under these conditions the rat can distinguish square from circle, there are two objections to Deutsch's theory which are of a more general character. The first has been mentioned already; the system allows of recognition of a shape in any orientation, but this does not appear to be a characteristic of primitive visual systems; the second is, that it does not allow for discrimination of part figures, or the isolation of a part of a *continuous* contour, although apparently it is just this sort of 'abstraction' that mediates shape discrimination in the rat. This would not be a serious objection for large figures, but the same process apparently occurs with small figures. Group I*b* (see above) treated several shapes with horizontal base lines as 'positive' although there is some evidence to suggest that when such figures are shown with a square the square is preferred (transfers (iii) and (iv)).

There follows a suggested system for shape recognition of the same logical type as Deutsch's but which accords better with the known facts—especially those mentioned in the preceding paragraph—and also satisfies the remaining four parts of Deutsch's specification (facts (*a*), (*c*), (*d*) and (*f*)).

IV. THE THEORY

Consider a two-dimensional array of units (or cells) arranged in chains, in such a way that each unit is connected to two neighbours, except for the end units of each chain, each of which has only one neighbour. Each chain is functionally separate from every other chain, except in two important respects: one set of end units (set A) is so arranged that if one fires, they all fire simultaneously, and the other set of end units (set B) discharges into a common channel, which, following Deutsch's terminology, is called the final common cable (f.c.c.) Each chain has the same number of units (this is not a necessary condition of the working of the system, but its introduction simplifies matters considerably). Each unit is connected to a receptor, or group of receptors, and when the receptor or group of receptors is stimulated this activates the unit to which it is connected. Such a unit is 'active'; when a unit is not so activated it is 'passive'. Adjacent receptors excite adjacent units. It is assumed, as in Deutsch's theory, that only excitation from regions of sharp change in stimulation (contours) arrives at the plane of units.

Any unit of the array can pass excitation from itself to a neighbouring unit in the same chain: this excitation originates in the firing of the end-set A, and is not to be confused with activation by stimulation of receptors to which the unit is connected. The passing of this excitation from one unit to another involves a small delay, the length of which is determined by whether the unit transmitting excitation is active or passive. Only the two states are envisaged, and the delays for each are constant,; say Ta and Tp. For convenience, suppose $Ta > Tp$.

The delay involved in excitation passing from one end of a chain of $n+1$ units to the other, when all units are passive, is nTp. The time taken to traverse any particular chain will depend on the numbers of active and passive units in the chain and can be written in the form $(n-r)Tp+rTa$. If all the units of the array are passive, the traverse times for all the chains will be equal, and there will be a single, large discharge in the f.c.c., since all the A set fire together. The signal in the f.c.c. will depend only on the number of active and passive units in individual chains, and not on the positions of those chains, or the positions of the units on the chains; this signal can be written simply as the sum of a number of terms of the type $(n-r)Tp+rTa$. Chains with equal numbers of active units will fire simultaneously into the f.c.c.

It is assumed that the array is so arranged that chains will represent, functionally, one of the principal axes in visual perception: thus stimulation of groups of receptors by, say, a horizontal line will fire units all on one chain, but which chain this is (i.e. the position of the horizontal line) will be immaterial as far as the signal in the f.c.c. is concerned, which will be a single large discharge (from all the 'empty' chains), followed after a delay by a single small discharge (from the one delayed chain). Such a system will be sensitive to small changes of direction; a small rotation from the horizontal will throw the activation off a single chain, and involve fewer units per chain on a *number* of chains. This means that the final signal now is a single large discharge (but smaller than in the previous case) followed after a shorter delay by a simultaneous discharge from several chains. The system is obviously most sensitive to changes of direction near the horizontal. The system will discriminate between simple geometrical figures, seen separately, such as a square and a triangle; both will give a signal consisting of three volleys: the first for the 'empty' chains; the second for the non-horizontal sides of the shapes, of high amplitude but short delay, caused by a few active units on a large number of chains; the third for the horizontal sides, of low amplitude but relatively much longer delay, caused by a large number of active units on a small number of chains. The main difference between the two signals is that the square's third discharge has about twice the amplitude of the triangle's, since the square has two sides in the horizontal and the triangle only one. Similarly, other simple shapes give rise to distinctive signals. However, as it stands the system would have difficulty in distinguishing between two rectilinear shapes neither of which had a side in the horizontal: the signals for a diamond and a pair of vertical lines, for instance, would be practically identical. This difficulty is overcome by assuming that there are two such systems in action, the second one functionally perpendicular to the first, i.e. a *vertical* line will stimulate receptors connected to units on a single chain; each system has a separate final common cable.

This double system has the interesting property that horizontal and vertical lines will be discriminated from one another much more readily than mutually perpendicular lines set at 45° to the horizontal. This effect has been observed in the octopus (Sutherland, 1957).

The system will account well for the experimental results reported above. One would expect shapes to be discriminated by presence/absence of a rectilinear contour lying in one of the principle axes if this cue is available in a particular case, since this gives a distinctive signal in the f.c.c., and that if this cue has been used as a basis for discrimination *any* figure having the property will be discriminated from any figure lacking it. The results reported above, though limited to a few shapes, are consistent with this.

In this particular case the pre-training figures were large, so the conditions favour isolation of an element of the shape. It remains to be seen if the effect persists under less favourable conditions of training.

The difficulty in discriminating small square from small circle (group I*b*, see above) could be explained either by supposing that a short line produces a smaller delay in the f.c.c. which is less distinctive than the long delay, or by supposing that with the smaller shapes more elements of the shape fall within the effective field of vision and hence introduce additional signals in the f.c.c. which tend to mask the delayed discharge on which discrimination is based. These are real alternatives, in the sense that an experiment could be set up to decide between them.

It is also perhaps worth mentioning that many of the results reported by Lashley can be accounted for by the system. To show this in detail for all Lashley's figures would require a separate paper, but two examples will illustrate the point.

(1) Rats trained to discriminate horizontal against vertical striations transferred to single horizontal against single vertical lines and interrupted horizontal against interrupted vertical; interrupted striations, horizontal against vertical, so long as the interruptions

were small compared with the width of the striations (Lashley, 1938, p. 147). It is of particular interest to note that there was no transfer to horizontal and vertical zigzags, and 'the same result was obtained in other tests with narrower lines and with wavy lines, *whenever the deviations from a straight line exceeded the width of the striations*' (1938, p. 148) (my italics). This is just the condition which would determine the disruption of the distinctive signal in the f.c.c.

(2) Comparison of two of the 'interference' situations shows that patterns which produce regular interference in the f.c.c. are more disruptive of a discrimination than those which produce 'random' interference. Thus the discrimination of triangle from cross is not disturbed by introducing a number of small isolated patterns around the discriminanda (Lashley, 1938, pp. 140–1, figures 20 and 21) but introduction of horizontal and vertical striations (figure 22) does disrupt a discrimination (see also 1938, p. 146). This would be expected: if one regards the f.c.c. as an information channel, the introduction of a certain amount of noise will interfere less with the reception of the signal than the superposition of a second signal on the first. (The introduction of a number of small isolated shapes into the display will generate something close to random noise in the f.c.c.).

It is not claimed that all Lashley's results can be so neatly explained by the theory, nor is this to be expected, since other factors besides shape recognition entered into his experiments.

The system has the advantage that it will explain equally well discrimination by part-figures or whole figures, and a discrimination need not break down in going from one to the other, provided the same distinctive signal (or part of a signal) is set up in the f.c.c. in both cases; in Deutsch's system the signal set up in the f.c.c. would alter radically even if only a small part of a figure fell outside the field of vision. It also avoids the somewhat artificial restriction, necessary for Deutsch's theory, of 'numbering' shapes if several are seen simultaneously (Deutsch, 1955, p. 35).

The limited scanning process required in this theory is free from the objections to such processes in general made by Deutsch (1955, p. 31).

Some of the predictions which can be made from the theory are as follows: (i) When *small* shapes are used, there should be confusion of upright with inverted triangle, but not with triangles in any other orientation, and in particular not with a triangle having one side vertical, although this shape would itself be confused with its mirror image. (ii) The prediction of lowered threshold for discrimination of the angle of inclination of straight lines near the horizontal has already been made: it applies equally to straight lines near the vertical. (iii) A discrimination would be upset, not only by rotation of the shapes, but also by rotation of the animal's receptors (e.g. by tilting the head to one side) but would be restored by rotating both the receptors and the shapes by equal amounts, and in the same direction. The testing of this prediction would present practical difficulties but it is possible in principle (Gellerman taught chimpanzees and children to discriminate a triangle from another shape; when shown a *rotated* triangle both rotated their heads so as to bring the triangle into the learned orientation (base-line horizontal) before making a discrimination (Gellerman, 1953, pp. 13–17).)

The theory fits the acceptable parts of Deutsch's specification (see above, §III), and also the characteristics of shape recognition found in the experiment reported in this paper. It has plausibility as a system which could be neurologically embodied; in this respect it presents less difficulties than does Deutsch's system, despite the fact that processes of spreading excitation in nervous structures of the type required for Deutsch's theory have been discovered (Burns & Grafstein, 1952; Motokawa, 1951). The theory is testable.

I should like to record my indebtedness to Dr J. A. Deutsch for his constant encouragement and advice, both during the course of the experiment and in the preparation of this paper.

REFERENCES

BOYCOTT, B. B., & YOUNG, J. Z. (1956). Reactions to shape in *Octopus vulgaris* Lamarck. (In the Press).

BURNS, B. D. & GRAFSTEIN, B. (1952). The function and structure of some neurones in the cat's cerebral cortex. *J. Physiol.* **118**, 412–33.

DEUTSCH, J. A. (1955). A theory of shape recognition. *Brit. J. Psychol.*, **46**, 30–7.

DODWELL, P. C. (1956). Some effects of pre-training on subsequent shape discrimination in rats. *Brit. J. Anim. Behav.* **4**, 81.

FIELDS, P. E. (1935). Studies in concept formation. II. A new multiple stimulus jumping apparatus for visual figure discrimination. *J. Comp. Psychol.* **20**, 183–203.

GELLERMAN, L. W. (1933). Form discrimination in chimpanzees and two-year-old children. I. Form (triangularity) *per se*. *J. Genet. Psychol.* **42**, 3–27.

LASHLEY, K. S. (1938). The mechanism of vision. XV. Preliminary studies of the rat's capacity for detailed vision. *J. Gen. Psychol.* **43**, 123–93.

MOTOKAWA (1951) Propagation of retinal induction. *J. Neurophysiol.* **15**, 339–51.

SUTHERLAND, N. S. (1957). Visual discrimination of orientation by Octopus. *Brit. J. Psychol.*, **48**, 55–71.

(*Manuscript received* 8 *July* 1956)

PUBLICATIONS RECENTLY RECEIVED

Behavior Theory and Conditioning. By K. W. SPENCE. Yale University Press; London: Oxford University Press. 1956. Pp. vii + 262. 36*s.*

This volume presents the Silliman lectures at Yale for 1956. Prof. Spence is the first psychologist to deliver these well-known scientific lectures, though following physiologists such as Sherrington (The Integrative Action of the Nervous System) and Granit (Receptors and Sensory Perception). It must have been no easy task to present psychology in this company and criticism from some quarters was inevitable. Prof. Spence has chosen to present in an introductory lecture a brief historical survey of psychology and then settle down in five further lectures to examine rat psychology and conditioning. This is a thoroughly workman-like effort—anything but a lecturer's *tour de force*—and the student with 'Spence' in one hand and 'Hilgard and Marquis' in the other can weigh the evidence in the field of conditioning for himself. With such a wealth of detail it would seem invidious to select individual points for criticism. Suffice it to say that Prof. Spence, as we might expect from so eminent a Hullian scholar, gives an authoritative presentation of the work of the behaviour theorists and at the same time concisely states the differences between Hull and himself. His readers should be able to make their own assessment of the position reached by this research for these lectures represent one of the clearest expositions which we have been given of behaviour theory.

In the last chapter on 'Complex Learning' Spence examines the situation where the rat has to choose between one of two bars in a double-type Skinner box and, to finish on a human note, there is a brief consideration of how paired-associate learning is influenced by anxiety levels. The author's approach is epitomized in his observations that though he is concerned with simple learning phenomena he suspects that 'as yet, we have hardly begun to uncover all of the complexities that underlie even such simple learning behaviour as is involved in these situations.' The reader who has tracked his way through derivatives of Hullian learning formulae and some fairly advanced mathematics may begin to feel restive but there are two points which he should consider. The behaviourists are attempting to derive and interrelate empirical laws from classical and instrumental conditioning. They believe, rightly or wrongly, that such laws would serve as a theoretical model to account for further complex learning. With such aims the behaviourist is right to attempt to understand the simple situation—it is no criticism of him that he finds it to be complex. Rather he is to be commended for his honesty. At the same time his industry and integrity is no guarantee that he is pursuing the right road. We can of course go east by going west. But psychologists who object to this course might examine their own alternatives. This is Prof. Spence's second point. He pleads—and how much the reviewer agrees with him—for psychologists 'to learn to make more effective use of theories instead of taking the negativistic attitude that theories are something to be proved wrong by fair means or foul'. For many of us who see the psychological civil wars as a period of furious and too often misdirected activity a more constructive and more tolerant phase seems long overdue. Creative research has no time to carp.

H. KAY

Interpersonal Diagnosis of Personality. By TIMOTHY LEARY. New York: The Ronald Press Co. 1957. Pp. xix + 518. $12.00.

In this important book Dr Leary describes a 'functional theory and methodology for personality evaluation' which has been developed and applied by the Kaiser Foundation Psychology Research Project, of which, latterly, he has been the director. In the Preface he describes the approach used as a dynamic behaviourism: 'There are two dynamic attributes. The first refers to the impact one person has or makes in interaction with others; the second refers to the interaction of psychological pressures among the different levels of personality. The behaviouristic attributes of the system derive from the procedure of viewing every response of the subject (overt, verbal, symbolic) as a unit of behaviour which is classified by objective methods and automatically sorted into the appropriate level of personality. The patterns and clusters of thousands of these responses, sorted into different levels, are then converted by mathematical techniques into indices and into a multilevel diagnostic code summary. They are then related to clinical events or prognoses.'

In the early chapters the author draws on a number of psychoanalytic sources, notably Sullivan and Erickson, and on a wide range of clinical experience, to formulate nine working principles as a basis for the method. It is stressed that the variables to be used are to be capable of measuring adjustive as well as maladjustive aspects of personality, that they are to permit the classification and measurement of data from different levels of behaviour, and that the final objective is to make possible predictions about interpersonal behaviour in a functional setting, in this case a psychiatric clinic.

Sixteen interpersonal dimensions of personality are used to classify the data. These are ordered on a circular two-dimensional continuum, in which a vertical axis represents dominance-submission, and a horizontal axis, hostility-affection.

Five levels of personality, defined according to the source of the data are described: Level 1, *Public Communication*, concerns the interpersonal impact of the subject upon others. Level 2 is represented by the subject's *Conscious Description* of himself and others in his world. Level 3 derives from the autistic projective fantasy productions: for example TAT, and is called *Preconscious Symbolization*. Level 4, the *Unexpressed Unconscious*, is defined by the subject's omission and avoidance of significant interpersonal themes. Level 5, the *Ego Ideal*, comprises the subject's statements about his interpersonal ideas, standards etc.

It is emphasized that data from any one level can only be understood in the light of data from other levels of behaviour. Illustrations are given, from clinical and other sources, to show the usefulness of the assessments obtained at each level, and of discrepancies between the levels, and between self-assessments and the assessment of others.

In the second half of the book the application of this approach to interpersonal diagnosis is further developed and methods of quantifying the data and of measuring the variability between levels are elaborated. Clinical and research data are reported in detail for eight personality types, which are categorized, on the basis of levels 1 and 2 information, in terms of their dominant interpersonal theme.

Finally some applications of the interpersonal system are described; for example to the diagnosis of hospitalized psychotics and of psychosomatic personality types in medical practice, to the analysis of group dynamics in an industrial management group, and to the prediction and the measurement of interpersonal dynamics in group psychotherapy. Valuable appendices give information about the construction and standardization of the instruments employed in the research, together with norms, conversion tables and weighted scores.

This brief sketch of some of the contents of the book cannot give any proper impression of its richness in ideas or of its breadth of concern. It is an exceedingly important contribution to the psychology of personality, and in particular to clinical psychology. In many ways the approach to personality study and to the research problems involved in it are in the tradition of H. A. Murray's 'Explorations'. The Kaiser Foundation Research may well have a little influence on clinical and social studies during the next decade.

Dr Leary makes it quite clear that he is well aware of the incompleteness of the achievement, particularly in the inability so far to integrate level 4 and level 5 data into the system. Many psychoanalysts will feel that the research has been concerned too much with the development of an ego psychology, and too little with the unconscious dynamics and the ways in which they determine the quality of interpersonal relations. Such shortcomings, however, should not take from the value of this pioneer work. Apart from the practical usefulness of the concepts and methods described, much has been achieved towards 'bridging the ancient and intolerable gap between the science of personality and the practice of psychiatry,' and towards the development of a unified theory of personality.

H. PHILLIPSON

Prediction Methods in Relation to Borstal Training. By H. Mannheim and L. T. Wilkins. London: H.M.S.O. 1955. Pp. viii + 276. 17*s.* 6*d.*

This volume represents 'the first piece of criminological research to be published under the auspices of the Home Office in accordance with the powers given by the Criminal Justice Act, 1948' and is virtually the first comprehensive criminal prediction study done in this country. It is the result of close co-operation between a criminologist and a social statistician supported by the penal authorities. The authors emphasize that 'the combined efforts of administrators and research disciplines' will be essential for further advances in this area. It is thus a landmark in many ways, a piece of research which cannot be ignored by future workers and which has already made considerable impact in Borstal and similar training fields. Perhaps there is even a danger that it will set the pattern of future investigators too firmly on a narrow basis of replication studies or too rigidly factual, analytic, statistical approaches.

The presentation is both scholarly and workmanlike. The introductory survey in chapter I of American and European prediction studies is exhaustive. Chapter III is a lucid and persuasive piece of teaching, explaining the basic ideas underlying prediction, its purposes and uses, even down to a warning that the major value of the study does not lie in the direct application of its experience tables to operational situations but in its demonstration of prediction techniques as research tools. The descriptive analyses and tables of chapter V provide a wealth of information linking all the conventional, mainly sociological data, with themselves and with outcome. One could quibble about some of the minor conclusions drawn here from time to time, especially, for example, where assumptions are made about the 'intentions' or the mental operations of people making ratings or judgements which are later codified. The individual case studies of chapter VIII are interesting but not particularly illuminating. The statistical work (chapters VI and VII and Appendices) is sophisticated and logically developed with efforts to enlighten rather than bewilder the less mathematically inclined. This is especially true of the discussion of the 'decision theory' approach and its implications. Cross-validation is included in the design. A final note (chapter IX) on the future of criminological research will be a source of realistic advice to future investigators on what sort of use can be made of the kinds of data at present in existence and what improvements are necessary here.

I should not regard the main 'practical' findings (that 'open' Borstal treatment is more effective than 'closed', and that length of detention beyond the mean of the study has little effect on success) as so firmly established as the authors claim. Their criticism also of 'subjective' judgements, while not entirely unjustified, appears to me to be unduly harsh. Their plea, however, for further 'operational research' on these and other matters has been convincingly made. This is no doubt an important book which has cleared a way and set a very high standard for future work.

R. L. MORRISON

The Study of Groups. By JOSEPHINE KLEIN. London: Routledge and Kegan Paul. 1956 Pp. ix+200. 21*s.*

Although this book is primarily intended for sociologists, it should be of interest to some psychologists, especially to those in applied fields. The material is presented in an attractive and straightforward fashion with very, very few wasted words—a fact that allows the author to bring together in a short book evidence from a large number of researches, ranging from those early rather crude laboratory studies of group strengths in task performances to the recent refined work by Bales, Bavelas and Leavitt and others on social structures and interaction processes. The researches are arranged according to their relevance to certain assumptions made in the text whereby more and more variables are successively allowed to contribute to the increasing complexity of group life as it emerges from the laboratory to real life. The chapters follow this arrangement also, culminating in an ambitious attempt at 'The Larger Society and Social Change'.

The theorist familiar with his bibliography of group studies will not find much new here (except for the Appendix, in which the author relates some of her findings to those of Moreno and Bion) but could regard this work as an excellent refresher. Those whose professional work involves them in group work for the purposes of diagnosis, teaching, or therapy, will undoubtedly find their insights the better from a study of this text, not the least for the clear summarizing statements at the end of each chapter.

G. M. STALKER

Discussions on Child Development. Vol. II. Edited by J. M. TANNER and B. INHELDER. London: Tavistock. 1957. Pp. 271. 28*s.*

This volume contains the proceedings of the Second Meeting of the World Health Organization Study Group on the Psychobiological Development of the Child, and continues the stimulating and diversified discussions of volume I (reviewed in this *Journal*, **48**, p. **76**). The broad theme of this volume is 'learning'. As before, one of the participants introduces a topic which is then thrown open for discussion. These topics are: electro-mechanical models of aspects of learning (Grey Walter); effects of perceptual deprivation in man and animals (Dalbir Bindra of McGill University); learning under stress in animals

(Howard Liddell); the cross-cultural study of learning of internalized standards of behaviour (John Whiting); and the development of abstraction in learning (Bärbel Inhelder): John Bowlby introduces a discussion of the film 'A Two-Year-Old Goes to Hospital', and Konrad Lorenz introduces a film illustrating some behaviour patterns of prematures, newborns, and infants up to about one year of age. As in volume I, the discussion is the important thing. Although this volume may perhaps sparkle less than its predecessor, it is to be recommended as thoroughly worthwhile reading for all psychologists. The constant interplay of queries and suggestions gives the reader the refreshing experience of watching creative thinking in progress, carries him over a wide range of topics, loosens up some of his own preconceptions, and leaves him with plenty to think about. It is one of the book's merits that most readers will want to jump into the discussion with comments of their own. I. M. L. HUNTER

The Psychology of Adolescence. By A. T. JERSILD. New York and London: Macmillan. 1957. Pp. xii+438. 35*s*.

At thirty-five shillings, this book is good value. It may be regarded as a sequel to the same author's well-known Child Psychology; it follows a similar plan and approach. The first five chapters deal with the physical and intellectual growth of the adolescent, often as viewed by the adolescent himself. The remaining twelve chapters deal with various aspects of emotional and social development. It is evident that it is in these aspects that the writer's real interests lie. He does not delve deeply into psychological theory, being content to treat his material at a more descriptive level. Not that the discussion is superficial; the documentation is thorough and up to date, the evidence is drawn from many sources and is allowed to speak for itself with a commentary by the writer, whose insight and maturity of judgement is frequently more illuminating than the evidence itself. This does not claim to be a profound or original study of adolescence. It is 'for those who are seeking to understand adolescents, written for students, teachers and parents...'. It is significant that it is about 'adolescents', not adolescence. Of its class, this book is as good as the best of them, and better than many. JAMES MAXWELL

The Psychology of Child and Youth Drawing. By HELGA ENG. London: Routledge and Kegan Paul. 1957. Pp. viii+205. 50*s*.

It is now twenty-six years since the publication of *The Psychology of Children's Drawings, from the First Stroke to the Coloured Drawing*. Unlike many children, Margaret continued to draw throughout adolescence, and in this book Dr Eng traces her artistic development from the ninth to the twenty-fourth year. The first ten chapters contain 114 drawings and paintings, with descriptive text, and the remaining eighteen summarize the work done in this much-explored field during the past seventy years. There is a full and useful bibliography.

One cannot help questioning some of Dr Eng's premises: for example, that children's drawing, uninfluenced by formal instruction, is necessarily 'free'; that subjects such as elves and fairies are evidence of imagination; and that copying is always conscious. It is just at the self-critical stage that Margaret has reached at the beginning of the book that, without suitable stimulation and guidance, copying is most likely to occur. This may account for the difficulty Dr Eng meets when she tries to relate the drawings to the theories of Herbert Read and Viktor Lowenfeld. The book contains at the same time too much and too little information. A full psychological study of Margaret, illustrated by her drawings, or an integrated account of work done in this field, would have provided a more useful contribution to an understanding of child art. JOYCE COOKSON

Psychology in the Classroom. By JAMES MAINWARING. London: University of London Press. 1956. Pp. x+118. 7*s*. 6*d*.

This book is addressed to students and teachers but one cannot help thinking that the latter group will know much of what it contains already and will have little interest in the rest!

The book is split into four sections. The sections on 'Learning' and 'Knowing at the moment', base their theoretical matter on a highly dubious model of Dr Mainwaring's own invention. In consequence, the practical illustrations in both sections are more readily digestible than are the abstract formulations.

The third section is limited to a consideration of a few of those problems associated with thinking. The aspects discussed are explained lucidly and in a more usual and acceptable theoretical framework. The final section on 'Feeling' is more philosophical than psychological, which perhaps explains why so little is said about individual differences. The old thorny problem of discipline, left to the last, rates a chapter to itself. This chapter provides a worthwhile text for any prospective primary teacher. Unfortunately it does little to help the poor harassed secondary teacher, who has to deal with children who had developed under a system far removed from the author's!

At times the terminology is muddling, as, for example, when the term 'physical response' is used to denote the sensory stimulus, and, throughout the book, far too little reference is made to the experimental aspects of educational psychology. Despite these limitations and the inadequacies on the theoretical side, this book should be very useful in introducing Training College students to the subject.

G. SHOUKSMITH

A Case of Spatial Inability in a Cerebral Palsied Child. By E. M. CALDWELL. London: British Council for the Welfare of Spastics. 1956. Pp. v + 24. 4*s.* 6*d.*

This booklet is written by a teacher for those concerned with the training of children suffering from cerebral palsy. The term 'spatial inability' is given to a syndrome which, in adults, is differentiated by neurologists into visuo-spatial disorientation, disorders of body-image, and the various types of agnosia, apraxia, aphasia, and acalculia. The disabilities are illustrated by an account of the progress of one child over a period of about 15 months and by a table giving brief notes of findings on thirty-three children aged from 6·6 to 13·4 in the Percy Hedley School at Newcastle. The various suggestions for special forms of instruction and for modification of the usual methods will be useful to teachers and therapists dealing with such children. It is to be hoped that the psychologist who provided the test results will be given space to elucidate and amplify his findings in the more detailed report promised for a later date.

J. A. L. NAUGHTON

Clinical Psychology: The Study of Persons. By RICHARD W. WALLEN. London: McGraw-Hill. 1956. Pp. 388. 45*s.*

This work is intended as a text-book for a first course in clinical psychology, and to be suitable for either advanced undergraduates, or the first year of a graduate course. A background of knowledge in the psychology of personality, normal and abnormal, and in psychological 'testing', is assumed. The author's theoretical position, he considers, 'leans toward neo-behaviorism': in practice, he adopts a reasonable eclecticism. Among the topics singled out for special consideration are: prediction in clinical psychology; clinical sensitivity and hypothesis making, with remarks on the blunting of sensitivity by some types of academic training, masquerading as scientific objectivity, and on excessive diffidence in the use of simple observation; interviewing; the use of tests, psychometric and projective, with special mention of Rorschach and T.A.T.; and psychotherapy. There is also a chapter on medical assessment, which attempts to cover too much in its space.

The book is naturally orientated towards the North American scene. In general, this is of importance only in relation to the practical illustrations, e.g., the line of inquiry which gets from 'dates' to 'clap' in five easy moves might strike the Briton as somewhat brusque. The necessary transpositions can readily be made by anyone interested in the subject.

The book is evidently the product of a teacher as well as a clinician, and takes this aspect of its function seriously. The text is frequently interspersed with problems related to the preceding subject-matter, rather to encourage speculation and an interpretive attitude, than just as a check on absorption-rate. The limitations of reading as an introduction to this type of work are, however, clearly recognized.

It is encouraging that a clinical psychology can be taught, which is not amateur psychiatry, simple psychometrics, nor merely a pretext for the use of clinical material to prosecute theoretical inquiries, and which is soundly based scientifically but keeps a firm grasp of practical exigencies.

J. A. L. NAUGHTON

Physiological Psychology. By M. A. WENGER, F. N. JONES and M. H. JONES. New York: Holt; London: Constable. 1956. Pp. viii + 472. 40s.

Intended as an introductory text, this work is, with some reservations, as good as one could reasonably expect to find in a field as broad as physiological psychology. Such books cannot be easy to write, if only because the authors are bound to lay themselves open to criticism regarding the selection of material. Certainly, in this particular case, it seems strange to find references to the work of, for example, Békésy, Galambos, Harlow, Magoun, Penfield and Selye, but no mention of, for example, Granit, Hunt, Lashley, Motokawa, Neff and Pribram. Given the interests of the senior author, it is understandable that the sections on autonomic functions and on psychosomatic inter-relations are better written than the remainder of the book (incidentally, is the term 'somatopsychics' really necessary?). The last chapter, in which the authors come to regard the preceding two dozen chapters as a 'survey of facts and principles', is supposed to highlight the issues currently important in research, but in fact has been limited in scope mainly to consideration of the subject-matter of the Milbank Memorial Fund's 1952 Symposium on the Biology of Mental Health and Disease. The book is adorned with two excellent diagrams from the Ciba Collection of Medical Illustrations, and a reasonably clear glossary of terms, but has an irritating reference system. 'Suggested readings' at the end of each chapter are supplemented by more extensive lists of sources at the end of the book, but these are arranged, again, chapter by chapter, and occasionally include such unhelpful entries as, for example, 'Wenger, M.A., Unpublished research' (p. 445)—this particular entry never being actually referred to in chapter 17. A few other entries, although referred to in the text, prove wrong on further inquiry (e.g. on p. 280 Ref. 8 should read Ref. 7). Misprints such as 'recticular formation' (p. 252) are inexcusable.

S. SZAFRAN

Acta Physiologica Scandinavica. Vol. 39, Supplement 134. Stockholm. 1956. Pp. 112. $3.00.

This volume comprises six papers. The first, by G. Svaetichin and R. Jonasson, is 'A technique for oscillographic recording of spectral response curves'. The remaining papers, by Svaetichin, are 'Spectral response curves from single cones', 'Receptor mechanisms for flicker and fusion', 'Notes on ERG analysis', 'The cone function related to the activity of retinal neurones', and 'Aspects on human photoreceptor mechanisms'. It is of particular interest that the authors have obtained spectral response curves by the stimulation of single cones in the retinae of Teleost fish in Venezuela. They used twenty-four almost monochromatic interference filters distributed over the visible spectrum as light sources. These fish have colour vision essentially the same as that of man. The experimental data give very strong support to a four-colour theory of colour vision not at all unlike that outlined by Houstoun in 1932, and there is strong evidence that brightness or subjective luminosity may be mediated in part by a special type of cone. Three types of response patterns were obtained from different cones, and not more than three. The first shows depolarization in the red region of the spectrum and hyperpolarization in the green; the second shows depolarization in the yellow and hyperpolarization in the blue; and the third shows depolarization in proportion to the luminosity of the whole spectrum. In Teleost fish, and in some diurnal vertebrates which have colour vision, some cones are large and single, while others are smaller and double. Svaetichin has suggested that the large single cones may give the response signalling luminosity, while the small double cones may be of two kinds, each kind giving alternative responses, one for red or yellow and the other for green or blue, according to the kind of double cone in action. No double cone can give both the red and green or both the yellow and blue signals at the same time (in normal colour vision). Svaetichin has suggested that the wave-lengths of the lights giving peak responses for each pair might be as follows for man, the maxima for fish being shifted 20 $m\mu$ to the short wave end of the spectrum: red, about 634 and 420 $m\mu$; green, about 486 $m\mu$; yellow about 590 $m\mu$; blue about 440 $m\mu$. The maximum for luminosity would be about 554 $m\mu$. These peaks for colour would correspond to the points of maximum hue discrimination for man. Achromatic subjects might lack all but the luminosity receptors.

If experiments such as Svaetichin and Jonasson have reported could be carried out on the human retina, and this is probably not an impossibility, much could be done to test four-colour theories, which appear to be strongly supported by observations on colour blindness, anomalous colour vision, and minor variations in colour sensitivity reported by various authors. Judging by experiments on human colour vision and its variability, the peaks for sepctral response curves for man would be about 671 mμ for red (with an unrevealed peak in the violet), 546 mμ for green, 593 mμ for yellow and 488 mμ for blue. These differ somewhat from the peaks suggested by Svaetichin, but the differences might be due to various factors, including the special ways in which the response curves combined in actual vision, and the effects of pigmentation of the transparent parts of the optical system.

R. W. PICKFORD

Research Films in Biology, Anthropology, Psychology and Medicine. By A. R. MICHAELIS. New York: Academic Press. 1955. Pp. xvi+490. $10.00.

Dr Michaelis, by training a chemist and by profession a consultant in scientific cinematography, has, we are told, 'made research films on numerous subjects, including the circulation of blood in frogs, the spinning of aeronautical models in wind-tunnels, and the behaviour of bubbles in liquids'. This is his guide book, specially written for those who would make or use scientific films of a biological, anthropological, psychological, or medical nature. It should be read by anyone contemplating using a ciné camera before, if possible, the camera is bought, since it will help him to make a wise purchase. To those who have a camera, it will suggest new uses and improved techniques: to those who but use films it supplies a background of knowledge which should help in the critical evaluation of what is presented.

Following a general introduction of 31 pages, Part I (129 pages) deals with 'The Biological Sciences'. Fifteen of these pages are devoted to animal behaviour. These include references to the work of Von Frisch, Lorenz, Köhler, Schlosberg and Katz, Liddell, Yerkes, Pavlov, Masserman and Mowrer. Part II (99 pages) is concerned with 'The Human Sciences', Anthropology, Psychology and Psychiatry. Twenty-five pages are devoted to Psychology, as here defined, with over a hundred references to films in Child, Educational, General and Industrial Psychology. Part III (101 pages) refers to 'the Medical Sciences'. Here we learn how the film has been used in the study of locomotion, digestion, respiration, etc. A final chapter deals with Aviation Medicine. Each part has its own special introduction devoted to specialized techniques. For example, endoscopic cinematography is discussed in the introduction to Part III. The book concludes with an excellent Author and Reference Index (91 pages with 1490 references) and a good Subject Index.

The division of the subject-matter implied by the titles to Parts I, II, and III is open to objection, especially since it may create the impression that this book is of less importance to the psychologist than in fact it is. I have called this a guide book, which it is, but I must add that it displays a degree of scientific detachment hardly characteristic of such books. Nothing is advertised: certain apparatus and methods are advocated. This book merits a place in every Psychology Department.

HENRY H. FERGUSON

Les Structures Rhythmiques. By PAUL FRAISSE. Publications Universitaires de Louvain. 1956. Pp. viii+124. No price given.

The neglect of this particular field of work during recent years is no doubt largely due to the difficulty of devising procedures which can effectively investigate rhythm. Prof. Fraisse is all the more to be congratulated for tackling the problem and for his very thorough analysis of the conditions and limits of rhythmic structure. The first experiments grew out of the work directed by Prof. Michotte at Louvain on the 'morphology' of motor reactions and the present book gives the results of twenty years work on the problem.

Periodicity and structure are both inherent in our experience of rhythm and there is a motor component which makes it an experience rather than a perception. The conditions influencing rhythmic grouping include: length of intervals, duration of sounds, accentuation, speed, and number of elements. Rhythmic structure has a temporal aspect, which assures the unity of the group, and an intensive aspect which appears only with the repetition of the group. The temporal structure turns out to be astonishingly consistent; the groupings make use of two intervals, a 'short' and a 'long', whose actual durations are of the order of 0·2+ sec. and 0·5+ sec.; the ratios between the various intervals in a group are mainly of the order of 1 and 2. Reproductions of groupings which do not make use of 'pregnant' forms involving simple and double ratios between intervals show the play of principles of 'assimilation' and 'distinction' which modify in the direction of pregnancy. Modifications tend to affect the 'long' interval, suggesting that the 'short' is more fundamental. It is not only the relationships within the group which are important; rhythm is not to be considered as repetition of isolated groups, but the experience depends on organization of successive stimuli which also takes into account the length of the pause between groups.

Looking for a final interpretation in terms of physiology, Fraisse distinguishes 'perception of collection' and 'perception of duration'; taking the duration of the perceptual process to be about 0·5 sec., perception of collection ('short intervals') occurs if intervals are less than 0·5 sec. and perception of duration ('long' intervals and pauses) occurs if the intervals are greater than 0·5 sec. 'Long' intervals and pauses are distinguished by being 'full' or 'empty' but the physiological basis of this distinction does not seem to be so well established.

Intensive structure is discussed briefly and the author himself admits that he has not offered very full support for his hypotheses. The evidence so far available, however, seems to justify the belief that temporal structure is the more fundamental.

The book concludes with a consideration of music and poetry in relation to the laboratory studies. The significance of the experimental results is clear and further analyses would be welcome in order to bridge fully the gap between the simplicity of the experimental rhythms and the complexity of music and language.

MARGARET R. PHEMISTER

The Psychology of Learning. By B. R. BUGELSKI. New York: Henry Holt; London: Methuen. 1956. Pp. xiii + 523. 50*s*.

There seem to be two obvious ways of planning a text-book on learning. Either we take it personality by personality and treat each theory as its proud parent would have us see it, or we can cover the ground topic by topic letting each contender have his say on each issue. Anyone nurtured on Hilgard will tend to overlook the fact that who-said-what-when is not necessarily still of interest when we come to assess the present status of learning theory.

As Bugelski, with what sounds like a tone of faint exasperation, puts it: 'The question of the variety or number of kinds of learning appears to be a pseudo-question'. This is a provocative thought, and typical of the panache with which he tackles the whole problem area. He picks his way through the tangle of conflicting viewpoints with an impressive sureness of foot. For the most part, it reads like a good lecture—the plan is not so obvious as to invite tedium, but everything seems to follow logically enough. The most recent literature has been combed with a good eye for the telling experiment and each reference has a way of popping up in just the right place.

Teachers who, since 1953, have wearied of the student whose standard plonking tactic is 'What about Osgood's mediation hypothesis?' will thank Bugelski for an account of mediated behaviour which is both clear and adequate. If there is to be a criticism of this book, it might be said that Bugelski is a little too fond of Hebb's brand of meta-neurology. Since Hebb has already warned us that he expects to find his views not more than 95% mistaken, Bugelski might have added a more cautionary note to the brilliant speculation. But, like it or lump it, there is little doubt that 'The Psychology of Learning' will be with us for the next few years.

R. T. GREEN

The Mentality of Apes. By WOLFGANG KÖHLER. Harmondsworth: Penguin. 1957. Pp. 286. 3*s*. 6*d*.

This inexpensive and very welcome volume is a reprinting of Köhler's classical monograph as it appeared in English in 1925. The nine photographic plates of the 1925 edition are omitted but, otherwise, nothing has been added or subtracted.

I. M. L. HUNTER

The Unconscious Motives of War. By ALIX STRACHEY. London: George Allen and Unwin. 1957. Pp. 283. 25*s*.

As would be expected from its author, this book contains a very clear and illuminating account of the basic ideas and themes of psycho-analysis. In the first part, ten chapters are devoted to the instincts, ego-functions and defences, to the origins of the moral sense, the stages of infantile development, dreams and regressive phenomena. It is useful that in the chapter on the perils of regression Mrs Strachey makes a serious attempt to deal with psychotic problems, because they are specially relevant to war. In the second part three chapters are devoted to the group influences underlying war. The social group is viewed as a regressive influence, at least in respect of its bearings on war, and the special dangers of the sovereign state as an institution eminently suited to act as a regressive group are well expressed.

In the third part various recommendations are made by which tensions leading to war might be reduced. These are divided into three groups, the first of which includes changes towards a reduction of authoritarianism and dictatorial control in the state. The second includes changes to be brought about in the individual by decreasing his frustrations and traumata in early infancy as much as possible, and thus establishing a healthily balanced attitude to adult life, not over-subject to group-proneness. The author does not underestimate the difficulties of reducing infantile traumata and frustrations. The last group of recommendations is about the value of psychoanalysis itself, and this leads to the concluding chapter, in which Mrs Strachey suggests that psychoanalysis might be applied with advantage to people in five categories as a prophylactic measure against war. These categories are: persons more than ordinarily hostile, aggressive and destructive; persons who have a marked instinct of domination; persons in positions of authority; persons who happen to be highly charged with State-mindedness; and those who are exceptionally group-prone.

Possibly the weakest aspect of this admirable book is that the author may underestimate the resistance which would be offered to her recommendations, valuable though they would be, especially because the major tensions leading to war are probably of a psychotic nature and less accessible to psychoanalytic amelioration than we could wish.

R. W. PICKFORD

Towards a measure of Man. By PAUL HALMOS. London: Routledge and Kegan Paul. 1957. Pp. viii+250. 28*s*.

This book consists of a string of essays in which the writer attempts to define the concepts of normality and abnormality in human behaviour and personality. Halmos attempts to educe 'an absolute measure of man', and he does so by relying, as he says, 'on metaphysical, biological, and psychological information'. It is no accident that these sources of information are listed in this particular order, and the reader should be warned that when Halmos says 'psychological' he in fact means 'psychoanalytical'.

Those who like speculative thinking, subjective arguments, and a freedom from the discipline which experimental facts enforce, will probably enjoy this book. Readers of a more empirical bent will not so readily acquiesce in the easy acceptance of Freudian dogma and the failure to survey the facts relevant to his thesis characteristic of Halmos's book. To take but one example, Halmos quotes with approval Bowlby's doctrine of mother-child separation on p. 63 of his book, but he fails to give a critical summary of the research literature which would show that this doctrine is extremely insecurely founded and probably erroneous. Halmos would probably argue that this was meant merely as an illustration, but a whole book built up on illustrations in disregard of established fact cannot be unreservedly welcomed.

In many ways it is a pity that the book is disfigured in these ways, because Halmos's objective is an important and a sensible one. A thorough review of the facts of adjustment and the conditions relating thereto would be of great value and could easily be fitted into the kind of philosophical framework Halmos has in mind. Such a book would have made a genuine contribution to science. As it is, *Towards a Measure of Man* is more likely to find its place in the category of Belles Lettres.

H. J EYSENCK

The Person in Psychology: Reality or Abstraction. By PAUL LAFITTE. London: Routledge and Kegan Paul. 1957. Pp. x+233. 25*s*.

The book—states Dr Lafitte in the Preface—is 'a general argument about the nature of psychology'. Psychology is said to be in a bad way. Its weaknesses are: 'insufficient facts, sharply conflicting interpretations, neglect of the study of the greater part of common experience and a wide gap between the person of professional psychology and the person as commonly known' (p. 4). All these will not be remedied by studying the person in an objective manner: the study of neurology and the study of animal behaviour are, for the most part, irrelevant to psychology at large; objective test measurements are hardly better.

To remedy its weaknesses, psychology should return 'to an open dependence on common experience' (p. 47). To the extent that social psychology does this, it is exempt from the severe strictures made upon other kinds of psychology. Psychological inquiry 'must include the detailed study of wide and complex areas of ordinary behaviour' (p. 100). When studying a person some 'short cuts' are available. One is 'substantial abstraction', so called because it seeks the substance of the person's recurrent behaviour; it is based on information ranging from 'ordinary sustained acquaintance' to various projective techniques. There are others called 'instrumental definition' and 'synthetic abstraction'; they refer to the ways of gathering and classifying data. A full investigation includes also the person's history, and his exceptional (non-recurrent) behaviour. All these methods of inquiry are amply illustrated by studies carried out by the author and by his colleagues at the University of Melbourne.

Methodological considerations loom large in the book, and the author attaches a great deal of importance to them. Comparisons are drawn between psychology and other disciplines. The nature of psychological laws is discussed at some length. Philosophers of science would probably find all this rather controversial ground.

The interest of psychologists in methodology is understandable. There has been a lot of discussion as to what psychology should be like. Such discussion implies a peculiar unity of the very many different kinds of studies that psychologists undertake. The nature or degree of this unity may be questioned. However, from the standpoint of the advancement of knowledge, discussions by psychologists of the merits and demerits of particular studies may well be more valuable than discussions of general methodology.

W. SLUCKIN

L'Instinct dans le Comportement des Animaux et de L'Homme. By various authors. Paris: Masson. 1956. Pp. 796. Paper 6.500 fr.; cloth 7.200 fr.

In 1954, the Fondation Singer-Polignac organized a symposium on 'Instinct'. The fourteen participants from the European continent included K. von Frisch, H. Hediger, O. Koehler and K. Lorenz; the American members were R. Gesell, D. Lehrman, C. Richter, T. Schneirla and (from Brazil) M. Autuori; the Britishers were J. B. S. Haldane, Mrs Haldane-Spurway and D. Morris. This substantial volume contains the twenty-one papers presented at the symposium (eleven in French, seven in English, three in German), along with transcripts of the discussions following each paper, an introduction by P. Grassé, and a postscript by Henri Piéron. The greater part of the volume is devoted to animal studies of the kind which have come to be known as ethology; but there are excursions into the realms of biochemistry and endocrinology, the physiology of the nervous system and of metabolism, and psychoanalysis and philosophy. It is impossible here, to list the titles of the papers, much less comment on them, but this is an important source-book for psychologists who are interested in animal behaviour and in the physiological and evolutionary aspects of their discipline. And for the theoretical psychologist it is a forceful reminder that there exists no easy answer to the question 'what is instinct?' The discussions make it clear that behaviour cannot be dichotomized into innate and acquired. As used in this symposium, 'instinct' is but a chapter heading in biology containing many subsections also called instinctive and still to be precisely distinguished from one another. But the richness of the research findings presented shows that ambiguity and even disagreement over the word 'instinct' need be no deterrent either to communication or to rewarding empirical investigation.

I. M. L. HUNTER

Lysergic Acid Diethylamide and Mescaline in Experimental Psychiatry. LOUIS CHOLDEN (Ed.). New York and London: Grune and Stratton. 1956. Pp. xi+85. $3.00.

A Drug-taker's Notes. By R. H. WARD. London: Gollancz. 1957. Pp. 222. 16*s*.

Research into the problem of psychosis is proceeding vigorously. The Symposium of the American Psychiatric Association, edited by Cholden, is perhaps the most significant publication to appear since the new discoveries of the Osmond, Smythies and Hoffer research team (1952, 1954). This book explores a theory: that the basis of the functional psychoses is biochemical; and relates to an accompanying research technique: the experimental production of altered mental states, 'model psychoses', by administration of hallucinogen drugs such as mescaline, lysergic acid diethylamide (LSD) and adrenochrome. Such a theory and technique logically imply the research collaboration of psychiatrist, psychologist and biochemist. In the Symposium, Rinkel (psychiatrist) discusses the Osmond–Smythies theory that some substance of the mescaline-LSD group is a necessary causal factor in functional psychosis, and the later Hoffer–Osmond–Smythies theory that adrenochrome may be this substance. Rothlin (pharmacologist), Director of the Sandoz Laboratory where the hallucinogenic properties of LSD were discovered, gives a comprehensive account of this drug's pharmacology. Among other contributions is Hoffer's (psychiatrist) account of experiments showing the inability of psychiatrist observers to discriminate between psychotic patients and model psychosis subjects to whom LSD has been administered, from sound-recorded interviews. Those who would stress social rather than chemical factors in the aetiology of psychosis will be interested in the contribution of Savage (psychiatrist) who, while accepting a biochemical causal process, discusses how to avoid mistakes in handling schizophrenics of a kind he believes are commonly made by the community and therapist alike.

Criticism must be made of the thinness of references to earlier investigators: Prentiss and Morgan, Havelock Ellis, Weir Mitchell, Beringer and Mayer Gross, Klüver, and others. For readers of this *Journal* a serious defect is the indifferent quality of the two papers which deal with the psychological side of mescaline and LSD reactions. In the first of these Sandison (psychiatrist) argues that hallucinogen drugs help in the understanding of neurosis as well as psychosis; this claim is advanced with a great deal more dogmatism than evidence. Sandison's account gains nothing from his use of concepts as diffuse as 'birth fantasies' and Jungian 'mandala symbols' when he might instead have chosen to refer to the careful work of Klüver on 'form constants' in both natural and experimentally produced hallucination.

The second of these two papers, by Aldous Huxley ('Man of letters') is of such poor quality that it is difficult to understand its inclusion. Huxley chooses repeatedly to ignore individually different reactions to mescaline. His concepts fall short of even minimal criteria of precision, and fanciful metaphors like 'the mind's antipodes' and 'the psychological equivalent of Australia', together with dubious generalizations on such themes as 'the almost universal passion for flowers' do not strengthen this as an objective communication on the psychological effects of mescaline. This contributor's over-generalizations about psychosis, mysticism and human nature in general on the basis of what appears to have been a single mesculine experiment on himself do not escape censure even within the Symposium itself. As Fisher remarks: 'I think that from Mr Huxley's mescaline experience there is 99 per cent Aldous Huxley and only one half gram mescaline' (p. 67). It is unfortunate for serious research in this field that his own mescaline experiment—about which Mr Huxley has already written two books—has been so much popularized by Mr Huxley and with the implicit assumption that everybody reacts in a standardized way; perhaps it is a pity that Huxley's inclusion in the Symposium has given another opportunity for these gross over-generalizations to be repeated. Yet it is to be hoped that this slim but important book will, taken as a whole, do something to attract both lay and specialist attention to the more major issues raised by the model psychosis experiment.

Mr Ward's account of a series of experiments with lysergic acid diethylamide on himself is a better book than its title, or the publisher's blurb on its dust cover, would suggest. It is well written, the six experiments in question being reported in a lucid and interesting way; Ward exhibits careful attention to reports by earlier investigators, insight in describing what happened to him personally, and caution about attributing his own experiences to everybody. The dust cover is unworthy of the book when it poses the question: 'Who is right, Mr Ward or Mr Huxley?' Since Mr Huxley took *mescaline* (on a single occasion) and Mr Ward took *lysergic acid* (in six experiments) it is difficult to attach much meaning to this question. If mescaline is to be mentioned at all then Mr Ward's careful observations on himself with LSD merit comparison with the best of the accounts of mescaline experiences, e.g. that of Weir Mitchell in 1896. Moreover, Mr Ward's attitude not only to the possibility of different individual reactions to the same drug, but also to different reactions of the same individual on different occasions is highly sophisticated and, in the reviewer's opinion, eminently sound. Because of this quality, and because he is a careful introspectionist who is also a good writer, Mr Ward's reports are of considerable interest to the psychology of the LSD reaction. Readers of this book will gain from it some understanding of the serious purpose behind model psychosis experiments which have continued since Prentiss and Morgan late last century, why they are proceeding vigorously at the present moment, and what it is all about.

PETER MCKELLAR

VOLUME 48 NOVEMBER 1957 PART 4

INDIVIDUAL DIFFERENCES IN INTENTIONAL AND INCIDENTAL LEARNING

BY MAVIS PLENDERLEITH AND LEO POSTMAN
Department of Psychology, University of California

This study investigates the role of linguistic habits in the intentional and incidental learning of verbal materials. Linguistic habits determine the differential responses made to the stimulus materials during practice. The stronger the differential responses evoked by the stimulus items the better they are recalled. Similarly, individual differences in amount of recall reflect variations in the strength (conventionality) of differential responses. These relationships are more pronounced when learning is incidental than when it is intentional, and also are stronger with nonsense than with meaningful learning materials.

I. INTRODUCTION

Selectivity is a pervasive characteristic of verbal learning. Whenever the task exceeds the subject's immediate memory span, only a fraction of the material can be fixated on any one trial. The order in which different parts of the material are retained is not haphazard but reflects the 'complex and subtle influence exercised by the behavior repertoire of the individual upon the fixation of new responses' (McGeoch, 1942, p. 304). Specifically, the transfer of earlier habits to a new task determines the subject's differential responses to the learning materials.

The importance of differential responses as mediators of learning has recently been stressed by several writers (Dollard & Miller, 1950; Mandler, 1954; Osgood, 1953). Differential responses are reactions which give identity and meaning to a stimulus. When the stimuli are verbal, differential responses are likely to be themselves verbal or symbolic. The responses may be explicit, for example, overt verbal associations, or implicit. Implicit responses may be conceived as fractional components of responses which originally had been explicit (cf. Osgood, 1953, pp. 397f.). A well-established implicit response may carry the meaning of a stimulus concurrently with other associations or in the absence of overt associations. In the memorization of verbal series, differential responses (whether explicit or implicit) may be assumed to play a dual part: (*a*) they are instrumental in the differentiation of the individual items, and (*b*) they serve to mediate the reproduction of the items at the time of the memory test. Since differential responses are well practised and familiar reactions, they are readily associated with the context of the learning situation and then can 'lead' the subject back to the learning items proper.

Verbal stimuli may vary widely with respect to the frequency with which they elicit differential responses in a sample of subjects (Glaze, 1928; Hull, 1933). The positive relationship between associative frequency (association value) and recall has long been recognized (McGeoch, 1942, pp. 157f.). Frequency of recall should also be a function of the *associative potency* of the learning items. The associative potency of an item refers to the strength of the differential responses which that item elicits (cf. Mandler, 1955). For a given verbal stimulus, degree of associative potency may be estimated from the amount of agreement in subjects' associative responses. The use of inter-subject agreement as a measure of associative potency is in accord with Marbe's Law, which states that the latency of associative responses varies inversely with the frequency of responses in the population (Woodworth & Schlosberg, 1954, pp. 61–5). Since latency may be used as an index of habit strength, degree of inter-subject agreement provides us with an estimate of the relative strength of differential responses to a given word.

The responses of individual subjects show different degrees of conformity to population norms. We shall refer to this dimension of individual variation as the *conventionality* of associative responses. Applying Marbe's Law as before, we assume that the degree of a subject's conventionality provides an estimate of the strength of his associative responses. The more conventional a subject's explicit differential responses to the learning items the better should be his retention of these items. The effect of individual differences in conventionality should, however, vary with the nature of the materials and the set of the learner. With items of high associative potency, the explicit differential responses given during practice may be of little consequence. Even if they are not made explicitly, conventional responses are likely to occur implicitly and will serve to mediate the later reproduction of the items. With items of low associative potency, on the other hand, the differentiation and subsequent reproduction of the learning materials may depend largely on the explicit differential responses made during practice. Thus, the conventionality of such responses should be more important in the learning of nonsense materials than of meaningful materials since the latter are apt to have a higher degree of associative potency.

As regards the set of the learner, the conventionality of explicit differential responses should be a more critical variable when learning is incidental than when it is intentional. Intentional learners will maximize whatever differential responses are available to them for purposes of differentiating and connecting the learning items. The motivation to learn may mask, at least in part, the effects of individual differences in verbal habits. When subjects are not motivated to learn, only strong differential responses will be elicited. Individual differences in conventionality should, therefore, have pronounced effects on the amount of incidental learning. An earlier study (Plenderleith & Postman, 1956) has already shown that incidental learning is more sensitive than intentional learning to individual differences in verbal and discriminative habits.

The present study tests the following hypotheses: (*a*) items of high associative potency will be recalled better than items of low associative potency; (*b*) the more conventional an individual's explicit differential responses the better will be his retention performance; (*c*) individual differences in conventionality will have more significant effects with items of low associative potency than with items of high associative potency; (*d*) individual differences in conventionality will have more significant effects under conditions of incidental than intentional learning.

II. THE EXPERIMENTAL PROCEDURE

(i) *Experimental design.* Two sets of learning materials—a series of nonsense syllables and a series of English words—were used. There were two groups of subjects, viz. one group of intentional learners and one group of incidental learners. Each group learned both sets of learning materials. Associative responses to the learning items were obtained during practice, and recall was tested immediately after practice by the method of free recall.

(ii) *Learning procedures and tests of association.* The first learning task consisted of a series of 20 nonsense syllables varying widely in association value according to Glaze's (1928) norms. There were five syllables each of 0, 33, 66, and 100% association value. One syllable of each of these association values appeared in each of five blocks of four syllables in the list. The syllables, printed in capital letters, were presented for 4 sec. on 2 × 2 lantern slides. Following each exposure the subjects wrote down a meaningful association suggested by the syllable. The associations were written in the appropriately numbered space of a record blank. A cardboard strip was used to cover all spaces except the one in which the subjects were writing. The subjects were given 6 sec. in which to record their association, after which the next slide was exposed. The intentional learners were, in addition, instructed to remember as many of the syllables as possible without regard to the order of presentation. No learning instructions were given to the incidental subjects.

A 5 min. test of free recall was given immediately after the presentation of the last syllable. The subjects were instructed to write down all the syllables they could remember. They were encouraged to write down guesses as well as items of which they felt sure. Incidental learners were asked whether they had attempted to learn the syllables or had expected a test of recall. Five subjects who gave evidence of having guessed the purpose of the experiment were replaced.

The second task consisted of the 100 stimulus words of the Kent–Rosanoff test of word association. The stimulus words were printed in columns of 25 each on two pages of a test booklet. For each stimulus word, the subjects were given 6 sec. in which to write down the first association that came to mind. The associations were written in a space next to the stimulus word. A cardboard strip was used to cover all stimulus words except the one to which the subjects were responding. The intentional subjects, but not the incidental subjects, were instructed to remember as many of the stimulus words as possible without regard to the order of presentation. A 10 min. test of free recall was given immediately after the association test. At the end of the recall test, incidental subjects were asked whether they had expected the recall test or had attempted to memorize the words. Only two subjects had to be replaced on the basis of their answers to these questions. (The failure of most subjects to guess the purpose of the second task was probably due to the use of the Kent–Rosanoff test. The subjects recognized this test as a procedure for personality assessment and hence did not anticipate the second test of recall.)

III. RESULTS

(i) *The association tests*. To test our experimental hypotheses, it was necessary (*a*) to divide both nonsense syllables and words into items of high and low associative potency, and (*b*) to measure the conventionality of subjects' responses to these items.

The associative potency of a verbal stimulus has been defined in terms of the amount of agreement in subjects' associative responses to that stimulus. For both the nonsense syllables and the words, sets of norms were available from which the degree of inter-subject agreement could be determined. For the nonsense syllables, the associative responses of 530 undergraduate students had been obtained in earlier investigations at the University of California. The frequencies of different associations were tabulated for each of the syllables, and the frequency of the most popular association was used as a measure of inter-subject agreement. On the basis of these measures, the syllables were divided into two groups of 10 each, one of high associative potency, and the other of low associative potency. For the syllables of high associative potency, the mean percentage of subjects giving the most popular response is 44·34, with a standard deviation of 12·26. For the syllable of low associative potency, the mean percentage of subjects giving the most popular response is 17·00, with a standard deviation of 8·83. In the case of the words, the recent college norms of the Kent–Rosanoff test (Russell & Jenkins, 1954) were used. The frequency of the most popular response was again used to measure inter-subject agreement. The words were divided into two groups of 50 each, one of high and one of low associative potency. For the words of high associative potency the mean percentage of subjects giving the most popular response is 53·90, with a standard deviation of 14·80, and for words of low associative potency the mean percentage is 20·97, with a standard deviation of 4·58.

The norms were also used to measure the conventionality of the associations given by the experimental subjects. For each association, the percentage of subjects in the standardizing groups giving the same response was determined. The mean of these percentages provides a measure of the average conventionality of a subject's associations. Separate measures of conventionality were determined for syllables of high and low associative potency, and for words of high and low associative potency. Thus, there were four measures of conventionality for each subject.

To gauge the stability of the measures of individual differences, odd-even reliabilities of the conventionality scores were computed for each group of items. The coefficients of reliability, corrected by means of the Spearman–Brown formula, are presented in Table 1.

Table 1. *Conventionality scores*

	Nonsense syllables						Meaningful words					
	High associative potency			Low associative potency			High associative potency			Low associative potency		
Condition of learning	Mean	S.D.	r_{tt}	Mean	S.D.	r_{tt}	Mean	S.D.	r_{tt}	Mean	S.D.	r_{tt}
Intentional	18·65	7·54	0·64	4·66	3·68	0·57	30·60	10·71	0·94	8·77	2·43	0·82
Incidental	17·08	8·07	0·51	4·24	2·26	0·36	33·01	8·60	0·85	9·12	2·66	0·80

Meaningful words give satisfactorily high coefficients of reliability. The coefficients are considerably lower for nonsense syllables. This difference between the two kinds of materials may be ascribed to the shorter length of the series of syllables—10 items in each subgroup as compared with 50 in the lists of words. When coefficients are estimated for series of syllables comparable in length to the list of words, the differences in reliability all but disappear. The low reliabilities militate, of course, against significant correlations between conventionality scores for syllables and recall. To the extent that such correlations are none the less obtained, they would provide *a fortiori* evidence for the strength of the relationship.

The means and standard deviations of the conventionality scores of the intentional and incidental subjects are presented in Table 1. The results for the two kinds of learners are quite similar and will be considered together. The scores of our experimental subjects conform to the normative classification of the stimulus items but at the same time reveal considerable individual differences among subjects. The conventionality scores for items of high associative potency exceed those for items of low associative potency by a considerable amount. At both levels of associative potency the conventionality scores are higher for meaningful words than for nonsense syllables. An analysis of variance of the conventionality scores is presented in Table 2. (For purposes of this analysis as well as subsequent analyses, the arcsine transformation (Snedecor, 1946, pp. 448–50) was applied to the conventionality scores in order to remove heterogeneity of variance.) While the intentional and incidental groups do not differ significantly, individual differences among subjects are a significant source of variance. Thus, the conventionality scores reliably measure individual differences in verbal dispositions. Both the associative potency of the items and the nature of the material (syllables versus words) are highly significant sources of variance. The significant interaction of these two variables shows that the difference between items of high and low associative potency is reliably smaller for nonsense syllables than meaningful words. The interaction of subjects and materials is significant, i.e. the differences among the subjects are a function of the stimulus materials. There is, on the other hand, no significant interaction of subjects and associative potency. For both sets of materials combined, the differences among subjects are independent of the associative potency of the items.

Table 2. *Analysis of variance of conventionality scores*

Source	D.F.	Mean square	*F*	*P*
Between subjects:				
Conditions of learning	1	0·48	0·01	—
Individuals	98	48·39	2·85	<0·01
Within subjects:				
Associative potency	1	21,511·76	1,310·89	<0·01
Materials	1	6,069·43	168·17	<0·01
Conditions × potency	1	2·34	0·14	—
Conditions × materials	1	95·55	2·65	—
Potency × materials	1	376·08	22·17	<0·01
Conditions × potency × materials	1	28·29	1·67	—
Subjects × potency	98	16·41	0·97	—
Subjects × materials	98	36·09	2·13	<0·01
Error	98	16·96	—	—

Table 3. *Mean number of items recalled*

	Nonsense syllables				Meaningful words			
	High associative potency		Low associative potency		High associative potency		Low associative potency	
Condition of learning	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Intentional	2·86	1·54	2·18	1·48	18·68	4·33	17·94	4·95
Incidental	2·88	1·55	1·90	1·39	17·78	5·62	17·18	5·00

Table 4. *Analyses of variance of recall scores*

		Nonsense syllables			Meaningful words		
Source	D.F.	Mean square	*F*	*P*	Mean square	*F*	*P*
Between subjects:							
Conditions of learning	1	0·85	0·32	—	34·45	0·85	—
Individuals	98	2·68	1·57	<0·05	40·48	3·87	<0·01
Within subjects:							
Associative potency	1	126·41	73·92	<0·01	22·45	2·15	—
Conditions × potency	1	0·24	0·14	—	0·24	0·02	—
Error	98	1·71	—	—	10·46	—	—

(ii) *The tests of recall.* Table 3 shows the results of the recall tests, and Table 4 presents analyses of variance of the recall scores. The level of recall for words is considerably higher than for nonsense syllables. The over-all mean percentage of words recalled is 35·79, whereas the mean percentage of syllables retained is 24·55. This difference is obtained even though the list of words was five times longer than the list of syllables. In general, the percentage of items recalled is inversely related to the length of the list (Shurrager, 1940), but the difference in the degree of meaningfulness more than overcomes the effects of the length of the series. Since degree of meaningfulness is correlated with associative potency, this finding is consistent with the hypothesis that increase in associative potency favours recall.

Consider next the effects of associative potency within the meaningful and nonsense series. Nonsense syllables of high associative potency are recalled significantly better than syllables of low associative potency. In the case of meaningful words, the difference between the two types of items is in the same direction but it is relatively small and not significant. Thus, intraserial variations in associative potency significantly influence the average recall of nonsense syllables but not of meaningful words. Clearly, associative potency is only one of the determinants of the amount of recall. Syllables of high associative potency are recalled less well than meaningful words of low associative potency. Meaningfulness favours recall, and degree of meaningfulness is correlated with, but not exhaustively defined by, associative potency.

Intentional learners recall more meaningful words and more nonsense syllables than incidental learners. The differences are, however, small and fall short of statistical significance. This result was not unexpected in view of our previous finding (Postman, Adams & Phillips, 1955) that the presence of highly familiar items reduces the difference between intentional and incidental learners. Moreover, the task performed by both intentional and incidental learners—recording a meaningful association to each of the stimulus items—is conducive to learning. Since our main interest is in the comparison of individual differences among intentional and incidental learners it was desirable to equate the two groups as far as possible with respect to the average amount of retention.

Table 5. *Coefficients of correlation between conventionality scores and recall scores*

	Nonsense syllables		Meaningful words	
Condition of learning	High associative potency	Low associative potency	High associative potency	Low associative potency
Intentional	0·14	0·17	0·09	0·23
Incidental	0·11	0·48*	−0·22	0·36*

* Significantly different from zero at the 0·01 level.

(iii) *Analysis of individual differences.* We now turn to an analysis of the relationship between the conventionality of associative responses to the stimulus items and recall. Table 5 shows the coefficients of correlations between the conventionality scores and amount of recall. There are only two coefficients which are significantly different from zero ($P < 0{\cdot}01$), viz. the correlations for *incidental* learners between conventionality scores on the one hand and recall of words and syllables of *low associative potency* on the other. We have assumed that the conventionality (strength) of explicit differential responses to the stimulus items should have a maximal influence on recall when the items are of low rather than high associative potency and learning is incidental rather than intentional. It appears that a combination of these two conditions is required to produce a significant relationship between conventionality and recall.

Even though the correlations for intentional learners are not significant, it is worth noting that with both types of material the coefficients are somewhat higher for items of low than high associative potency. It is also of interest that with one exception (intentional learners with items of low associative potency) the positive correlations are higher with nonsense syllables than with meaningful words.

The pattern of correlations supports the conclusion that the relationship between the conventionality of differential responses and recall is a function of (*a*) the meaningfulness of the materials, (*b*) the associative potency of the items, and (*c*) the set of the learner. The relationship is maximized when the items lack standard meaning, have low associative potency and there is no set to learn. Any deviation from these optimal conditions reduces the magnitude of the relationship.

IV. DISCUSSION

The experimental results support the conclusion that the strength of the differential responses evoked by stimulus items significantly influences recall. Two lines of evidence have been presented: (*a*) items of high associative potency are recalled better than items of low associative potency, and (*b*) the conventionality of explicit differential responses given during training is positively correlated with recall. At the same time, our findings point to important limitations upon these generalizations. Associative potency has significant effects on recall with nonsense syllables but with meaningful materials the effects are small and not significant. Degree of meaningfulness is in part a matter of associative potency but clearly these two variables cannot be considered equivalent to each other. Quite apart from the strength of associative responses, meaningfulness has also been related to the number of different associations evoked by a stimulus item (Noble, 1952*a*; Mandler, 1955). This property of verbal stimuli has been shown to influence recall (Noble, 1952*b*), but is not taken into account in our measure of associative potency. We conclude that associative potency is only one of the characteristics of verbal stimuli which determines selective recall.

Similar limitations apply to the relationship between the conventionality of explicit differential responses and recall. Here significant correlations were found only with items of low associative potency under incidental conditions. As we have suggested, the associative context of high-potency items may be aroused implicitly and mediate recall even in the absence of explicit differential responses. The fact that significant correlations are found under incidental but not under intentional conditions makes it clear that motivational factors can maximize the effectiveness of relatively weak verbal habits in mediating recall. This interpretation receives general support from earlier findings that incidental learners show greater selectivity for familiar items than do intentional learners (Postman *et al.* 1955; Postman, Adams & Bohm, 1956).

Our experimental results dovetail with the findings of other experiments which point, directly or indirectly, to the importance of differential responses as mediators of recall. Hovland & Kurtz (1952) have shown that preliminary familiarization with a group of nonsense syllables facilitates subsequent serial learning of these syllables. It is reasonable to assume that at least one of the effects of the preliminary training was to attach differential responses to the syllables which served to differentiate them and to reinstate them in recall. Kurtz & Hovland (1953) have also shown that explicit naming responses at the time of training have significant beneficial effects on the recall of a series of objects. Here the role of differential responses in mediating recall is directly in evidence.

Our results point to certain methodological conclusions: (*a*) the determinants of selectivity in verbal learning are most readily accessible to experimental investigation under incidental conditions; (*b*) the role of verbal dispositions in mediating recall can be brought out most clearly in the learning of nonsense series or words of a low associative potency; (*c*) memory performance can be controlled by preliminary manipulation of the

Bureau Ednl. Psy. Research
DAVID HA... TRAINING COLLEGE
Dated ....

differential responses evoked by the learning items; (*d*) individual differences in selective retention can be related to the specific verbal dispositions with which the learner enters the experimental situation.

REFERENCES

Dollard, J. & Miller, N. E. (1950). *Personality and Psychotherapy*. New York: McGraw-Hill.

Glaze, J. A. (1928). The association value of non-sense syllables. *J. Genet. Psychol.* **35**, 255–67.

Hovland, C. I. & Kurtz, K. H. (1952). Experimental studies in rote-learning theory: X. Prelearning syllable familiarization and the length–difficulty relationship. *J. Exp. Psychol.* **44**, 31–40.

Hull, C. L. (1933). The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.* **45**, 730–4.

Kurtz, K. H. & Hovland, C. I. (1953). The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *J. Exp. Psychol.* **45**, 157–64.

McGeoch, J. A. (1942). *The Psychology of Human Learning*. New York: Longmans Green.

Mandler, G. (1954). Response factors in human learning. *Psychol. Rev.* **61**, 235–44.

Mandler, G. (1955). Associative frequency and associative prepotency as a measure of response to nonsense syllables. *Amer. J. Psychol.* **68**, 662–5.

Noble, C. E. (1952*a*). An analysis of meaning. *Psychol. Rev.* **59**, 421–30.

Noble, C. E. (1952*b*). The role of stimulus meaning (*m*) in serial verbal learning. *J. Exp. Psychol.* **43**, 437–66.

Osgood, C. E. (1953). *Method and Theory in Experimental Psychology*. New York: Oxford University Press.

Plenderleith, M. & Postman, L. (1956). Discriminative and verbal habits in incidental learning. *Amer. J. Psychol.* **69**, 236–43.

Postman, L., Adams, P. A. & Bohm, A. M. (1956). Studies in incidental learning: V. Recall for order and associative clustering. *J. Exp. Psychol.* **51**, 334–42.

Postman, L., Adams, P. A. & Phillips, L. (1955). Studies in incidental learning: II. The effects of association value on the method of testing. *J. Exp. Psychol.* **49**, 1–10.

Russell, W. A. & Jenkins, J. J. (1954). *Studies of the Role of Language Behavior*. University of Minnesota, Tech. Rep. no. 11.

Shurrager, H. C. (1940). The measurement of memory on an absolute scale. *Psychol. Monogr.* **52**, no. 234, 21–38.

Snedecor, G. W. (1946). *Statistical Methods*. Ames: Iowa State College Press.

Woodworth, R. S. & Schlosberg, H. (1954). *Experimental Psychology*. New York: Holt.

(*Manuscript received* 6 *September* 1956)

RUBELLA DEAF CHILDREN: A RORSCHACH STUDY EMPLOYING MUNROE INSPECTION TECHNIQUE*

By D. MARJORIE BINDON

Department of Psychology, Canterbury University College, N.Z.

With procedure altered to eliminate the deaf subjects' language handicap, the Rorschach test treated by the Munroe Inspection Technique was given to three matched groups of 15-year-old children: (*a*) thirty-six rubella deaf, (*b*) fifteen born deaf through causes other than maternal rubella, or who had gone deaf before the age of two, and (*c*) thirty non-deaf. All quantitative and qualitative results for both deaf groups were strikingly similar: neither cause of deafness, degree of deafness nor the residential factor was found to have had any collateral influence on their performances ($P > 0{\cdot}1$ in all cases). A criterion score of > 2 statistically selected Munroe checklist entries differentiated both deaf groups from the hearing group ($P < 0{\cdot}001$ in both cases). It is contended that the rigid personality pattern found to be common to the deaf is not abnormal, but subnormal (immature or undeveloped). It is contended also that their arrested personality maturation is relatively permanent and is not so much due to deafness itself, as to its consequences—a retardation in normal language development which has hindered the process of socialization, and hence full personality development.

I. Introduction

The present investigation is an attempt to study the personality of deaf children. The Rorschach test was used to reduce to a minimum the language handicap of the deaf subjects; to this end its procedure was also varied. Inter-personal communication was minimized by having no trial blot or initial explanation of the ink-blots, and by holding the short inquiry after each card and determining only the location category for each response. The test was administered individually, but the results were treated by the Munroe Inspection Technique for group administered protocols.

Past Rorschach studies of deaf children have been undertaken by psychologists not qualified to deal with the deaf, and also have serious flaws in experimental design: very small numbers used at each age level, differences noted not tested for statistical significance, no control group, or failure to control R (e.g. McAndrew, 1948*a*, whose average R for his deaf group was 9·0, for hearing group, 30·7), comparison of results with traditional adult norms. Hence the differing findings of previous Rorschach studies (Altable, 1947; Levine, 1948; McAndrew, 1948*a*, *b*) appear in general unreliable, and the present results will therefore not be related to them.

Although rubella deaf children have not been the subject of any published personality research, it is assumed in New Zealand (and overseas, apparently, e.g. Murphy, 1952, p. 119) that they behave differently from other deaf children. From her experience the writer formed the impression that this was not so, but at the same time, that deaf children in general differ from the hearing. Hence, besides the more general aim of attempting to contribute to our knowledge of the personality of the deaf, the present Rorschach study aims to test the following specific hypothesis: that rubella deaf children and

---

* Based on part of thesis in partial fulfilment of requirements for the degree of Master of Arts in Psychology, University of New Zealand. Thanks are due to Dr T. H. Scott, Canterbury College, for his helpful criticisms of the manuscript.

'non-rubella deaf children' (deaf before the age of two due to causes other than maternal rubella) display a *common* personality pattern, and that they are less mature than their hearing counterparts.

II. PROCEDURE

(i) *Subjects.* (1) Experimental group: thirty-six 15-year-old rubella deaf children. (2) Non-deaf group: thirty hearing children. (3) Non-rubella deaf group: fifteen children who were born deaf through causes other than maternal rubella or who had become deaf before the age of two. The two control groups were paired for sex, educational stream, age, and socio-economic status as gauged by parental occupation, with individuals from the experimental group. An additional criterion of degree of hearing loss was also used to equate the two deaf groups.

None of the deaf subjects had naturally acquired speech and language (hearing loss >60 decibels), and all were still attending government schools for the deaf in New Zealand where only the oral system of education was in use. The schools are all co-educational and admit both day pupils and boarders.

(ii) *Administration.* The Rorschach test followed a performance intelligence test, and was administered with S facing E (the writer) to facilitate speech-reading. The following specific instructions were given orally. 'This test I am going to give you is quite different from the first one. There are no right answers and no wrong answers. You can have as long as you like. In this card here [Card I held up, back view facing S] many people see all different kinds of things. I want you to tell me all you can see. Do you understand?...Right. Tell me when you are finished.' [Card handed to S.]

The deaf S's were free to supplement their verbal responses with their own sign language, drawing and writing. If S gave just one Rorschach response he was told for the first card only: 'Some people can see more than one thing. You saw a —. Can you see anything else? If you cannot that's all right because there are no right answers and no wrong answers.'

This instruction was included in case a deaf S might have misunderstood the opening test instructions. The inquiry was held after each card and determined only the location category for each response; it provided also a check against misnaming, or inaccurate recording of the deaf's responses by E. After the inquiry, the first card was withdrawn and the second was handed to S without comment. If S asked whether he could twist the cards he was told he could do as he wished.

(iii) *Scoring.* Though the test was administered individually, it was scored according to the Munroe Inspection Technique instructions for group-administered protocols (Munroe, 1950, pp. 117–34). The total composite checklist scores included both the checklist entries proper and the entries in the checklist left-hand column. The Munroe checklist is based on the Klopfer & Kelley system of scoring, but E deviated from their policy of scoring M (1946, pp. 109–13) and followed Rapaport (1949, p. 209) because the severe language disability of the deaf could otherwise prejudice their chances of scoring M responses by making them more prone than the hearing to assign only a name to what they saw in the ink-blots.

III. RESULTS

(i) *Quantitative results*

To avoid capitalizing on possible isolated large chance differences (cf. McNemar, 1949, p. 245) groups were not compared on their individual checklist category scores unless, using χ^2, their Munroe composite checklist score differences yielded a P value of less than 0·01. (χ^2 corrected for continuity, D.F. = 1, cutting point lower median score.)

(a) *Munroe composite checklist scores*

The Munroe composite checklist scores did not significantly differentiate between either the two matched deaf groups, or equated deaf subgroups formed on the basis of degree of deafness, sex and the residential factor respectively (see Table 1). That the Rorschach test did not discriminate between the deaf population supports the writer's view of no behaviour difference existing between rubella and non-rubella deaf children.

On the other hand, in accordance with the writer's hypothesis that deaf children in general differ from non-deaf, the Munroe composite checklist score *did significantly differentiate* the rubella deaf from its matched hearing group ($P<0{\cdot}001$, see Table 1).

(b) Comparison of rubella deaf and non-deaf in individual checklist categories

In view of the above result, the scores were examined for group differences in the separate checklist categories (see Table 2 for summary of findings).

Table 1. *Frequency and χ^2 of Munroe checklist composite scores for matched groups using lower median score as cutting point*

Group	No. of pairs	Frequency	χ^2	*P*
Matched rubella deaf group Non-deaf group	30	27 14	11·09	<0·001
Matched rubella deaf group Non-rubella deaf group	15	7 8	—	—
Rubella deaf boys Rubella deaf girls	16	12 8	0·602	>0·3
Partially deaf rubella* Total deaf rubella	18	8 8	—	—
Residential deaf group Non-residential deaf group	12	6 6	—	—

* Arbitrarily defined at those rubella deaf with average hearing loss >60 decibels but <101 decibels in the better ear as tested by a pure tone audiometer in the frequencies 512, 1024 and 2048. Their residual hearing could be utilized to an extent with the help of an individual hearing aid; hence they have acquired language at a faster pace than the totally deaf rubella (term arbitrarily designating the rubella deaf with average hearing loss >100 decibels) and consequently are in separate classes from the totally deaf rubella of similar age and intellectual ability.

Table 2. *Frequency and χ^2 of certain Munroe checklist entries for matched rubella deaf group and non-deaf group*

Checklist entry ...	−				− −				− − −			
Category...	Rub. deaf	Non-deaf	χ^2	*P*	Rub. deaf	Non-deaf	χ^2	*P*	Rub. deaf	Non-deaf	χ^2	*P*
P, Com	21	9	8·06	<0·01	10	2	5·10	<0·05	—	—	—	—
Range	28	21	4·01	<0·05	19	5	11·74	<0·001	16	2	13·41	<0·001
Fk, Fc	19	7	8·21	<0·01	18	3	14·36	<0·001	—	—	—	—
M	27	20	3·54	>0·05	25	13	8·68	<0·01	—	—	—	—
Total Mvt	28	15	11·82	<0·001	25	9	15·27	<0·001	20	4	15·63	<0·001
FC	25	12	10·15	<0·01	20	7	9·70	<0·01	—	—	—	—
	+				+ +				+ + +			
Dd %	27	23	1·08	>0·2	17	6	7·05	<0·01	—	—	—	—
F %	30	28			27	11	16·15	<0·001	18	2	16·88	<0·001
	B or *B/V*											
W	12	0	12·60	<0·001								
O	13	0	14·14	<0·001								
F	14	0	15·75	<0·001								
1 or more *B* or *B/V*	17	0	21·01	<0·001								

(c) *Generality of certain Munroe checklist entries*

To confirm in the deaf the presence of a distinctive behavioural pattern *common to the majority of its population*, the following steps were taken. (1) Eight checklist entries, henceforth called 'signs', were selected from Table 2: *Dd*++, *P*, *Com*−, *Range*−−−, *F*+++, *Fc*−−, *Total Mvt*−−−, *FC*−−, and one or more *B* or *B/V* entry: each represents a different checklist category and statistically differentiates the matched rubella deaf group from the non-deaf group ($P<0·01$). (2) These signs were then applied to individuals in the rubella and hearing group. Compared with four (of thirty) non-deaf, twenty-four (of thirty) rubella deaf children scored >2 signs, this difference being highly significant

($\chi^2 = 24{\cdot}17$, $P < 0{\cdot}001$). (3) To check further the differentiating power of these signs, they were applied to the non-rubella deaf group where twelve out of fifteen gained > 2 signs. This criterion significantly differentiated them from the hearing group ($\chi^2 = 17{\cdot}81$, $P < 0{\cdot}001$). Thus whereas 80% of the children in both deaf groups obtained > 2 signs, only 13·3% of the hearing did so. Quantitative results therefore suggest that there is a distinctive behavioural pattern common to the deaf children tested (rubella and non-rubella). Furthermore, the individual sign distribution within the two deaf groups was very similar and contrasted markedly with that of the hearing sample (cf. Fig. 1).

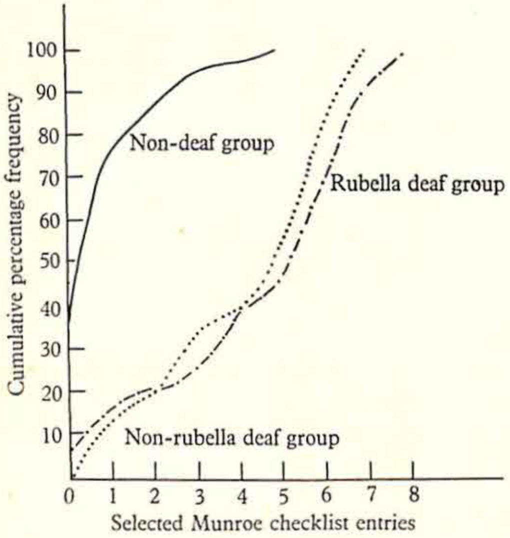


Fig. 1. Cumulative percentage frequency curves for rubella deaf group, non-rubella deaf group and hearing group showing their percentage distribution on eight selected Munroe checklist entries.

(ii) *Qualitative findings*

This section will discuss only that qualitative information which bears out the described quantitative differences. (The qualitative characteristics of the rubella deaf are those also of the non-rubella deaf.)

(*a*) *Form level quality of rubella deaf and hearing Rorschach responses*

It was only in the form level quality of the Rorschach responses that the rubella deaf and non-deaf children differed in direction of deviation from the Inspection Rorschach Record 'norms' of zero; all other group differences being in degree of deviation (cf. Table 2). As will be seen later, it is this form level quality of responses that supplies the key, not only to the characteristic Rorschach performances of the rubella-deaf population, but also to those of the hearing children tested. It should be noted first, however, that the checklist data alone (*B* or *B*/*V* entry, seventeen rubella deaf, no non-deaf) do not indicate the full extent of this difference in the form level quality of responses: all but four rubella deaf, as against no hearing children, gave $F-$ responses. Further, the deaf protocols usually revealed an indiscriminate mingling of $F-$, F and $F+$ responses, and though their productivity was high (av. R 30·2, non-deaf av. R 35·2), it was achieved at the cost of a decrease in quality. The finding that none of the non-deaf group produced $F-$ responses points up clearly the outstanding feature of the hearing protocols—their accurate form dominated quality. This 'respect for reality' on the part of the non-deaf revealed itself on the Rorschach records in several ways, for example: cards rejected; *W* responses displaced by *D* responses; the *Dd* category swelled with *dr* responses (which were not so much responses to very small aspects of the ink-blots, as responses involving the supplementing of *D* or *d* responses with parts of the adjacent areas); vague non-committal responses, rated as *F*, given to those parts of the blots which were themselves indefinite in form. (It is of interest to note that other Rorschach studies of New Zealand children—Malcolm (1951), Spinley (1949), Stroobant (1951)—all found average F% to be much higher than the F% norms for American children.)

It has already been mentioned that the deaf population did not have this regard for accuracy. Although, like the non-deaf, they too (1) rejected cards, (2) displaced *W* responses by *D* responses, (3) gave an excess of *Dd* responses (which for the deaf, however, were usually responses to very small blot areas (*dd*), and (4) gave vague responses, the explanation of their results differs. It is suggested in the next section that these types of responses are, for the deaf, those which can be given with the minimum of intellectual and perceptual effort and organization.

(*b*) *Certain determinant and content category responses of the rubella deaf*

F − responses. In order of frequency these were mainly (1) *DW* and other confabulatory responses, (2) inaccurate outline responses (due usually to assigning a specific concept to an indefinite form area), and (3) arbitrary responses (generally *Po* responses, or definite form responses drawn in on a formless area; the extreme example here being those few responses drawn in on the white part of the Rorschach card outside and independent of the blot proper).

F responses. Usually evasive form responses such as 'maps' and vague anatomy responses (often small *dr* responses).

F + responses. The majority were *dd* responses (e.g. foot, mouth, eye, tail). Edge detail responses were next in frequency. Unlike the majority of the hearing group's *de* responses (profile responses), their *de* responses were outlines of such parts of the body as shoulder, waist, mouth and top of head. The many *Do* responses also came into this category.

Content category. Although taken together the rubella deaf responses covered a wide range of associations, their content was predominantly a list of small parts of the body of either animals or humans, anatomy responses, and human object responses (e.g. shoe, belt, pocket, jersey). The hearing group, too, gave many *Hd* and *Ad* responses but they were to larger portions of the blot area and also represented larger portions of the body. Again, the thematic perseveration seen in some of the deaf protocols was rarely present in the non-deaf Rorschach records.

(c) *Handling of Rorschach cards*

In the actual handling of the Rorschach cards another important difference between the deaf and hearing subjects was apparent. The non-deaf children handled the cards in a normal manner but the behaviour of many in the deaf group resembled that of individuals of abnormal populations (e.g. purposeless twisting of the cards, 'edging' behaviour which Beck (1946, p. 60) reports occurs exclusively in schizophrenics, tracing outlines of blot areas with their fingers, and, more commonly, 'lapping' the cards which Ames (1952, p. 284) considered to be at least suspect behaviour even for children between two and ten). This last often produced *dr* responses which, though in a sense arbitrary, could not be considered *F −* because the blot area selected by this procedure was usually consonant with the configuration of the concept chosen.

IV. DISCUSSION

(i) *Personality pattern of the deaf—abnormal or subnormal?*

The characteristic personality pattern of the deaf with its marked overemphasis on pure form responses and unusual detail responses, very few popular, common, shading and form colour responses, extremely high number of responses confined to the human and animal detail category, poor form level, and a total absence of human, animal, or inanimate movement, resembles those described for such abnormal groups as clinic children (Beck, 1931; Salfield, 1950; Suarès, 1938), neurotics (Beck, 1937; Rapaport, 1949; Mons, 1948, pp. 118, 119), organic brain cases (Klopfer & Kelley, 1946, ch. XIV) and schizophrenics (Beck, 1937; Rapaport, 1949; Rickers-Ovsiankina, 1938). This finding, however, is not necessarily evidence for the presence of psychopathology in the deaf since their personality configuration also resembles those of young normal children (Ames *et al.*, 1952; Ford, 1946), institutionalized children (Goldfarb, 1944), normal adults under laboratory stress—produced situations (Broida & Thompson, 1954; Cox and Sarason, 1954; Eichler, 1951; Siipola & Taylor, 1952), and mental defectives (Beck, 1937; Klopfer & Kelley, 1946, ch. XVI).

This similarity of Rorschach results for such widely differing groups is not due, apparently, to identical thought content, but rather to their similar approach to the handling of the test stimuli brought about, it seems, by a general factor common to them all—the factor of restriction. This factor implies a lack of variability and adaptability and hence implies the presence of rigidity. Where this rigidity is accompanied by a disorganization of functions it is an indicator of abnormality; but where its characteristic feature is a lack of differentiation of functions it is an indicator of subnormality (immaturity or non-adulthood) and is a basic characteristic of undeveloped organisms (see Werner, 1940, 1946).

Since a person's thinking bears the stamp of his personality organization (Mayman, Schafer & Rapaport, 1952), it is the concept formation aspect of the Rorschach test which, together with a knowledge of the past history and present functioning of a specific group, will help to decide whether the rigid behaviour of that group is caused by functional disorganization or lack of differentiation. An analysis of the deaf's conceptual responses shows that, like younger normal children and mental defectives, they have not reached full conceptual development (cf. next subsection). They have been retarded by the restrictive factor of deafness which, in preventing normal language development, has seriously impeded their progressive socialization—a condition which is essential for ensuring full contact with reality and reality testing, and hence for complete cognitive development (cf. Piaget, 1951). The present Rorschach study reveals, therefore, that the restrictive personality configuration of the deaf may be termed a subnormal pattern and indicates that, in comparison with their hearing counterparts, they are functioning on a less mature level of personality integration.

(ii) *Rorschach results in relation to the conceptual processes of the deaf*

The characteristic Rorschach performances of the deaf listed below all required a minimum of intellectual effort and organization and are what the developmentalists would call 'genetically low' responses (Phillips & Framo, 1954): (1) Crude form responses represented fully on the cards (together with an absence of shading responses, where neither texture nor depth is actually on the cards but must be supplied by the subjects' imagination (Thomas, 1955), and a total lack of movement responses which are solely a product of the subjects' creative powers and may be termed the genetically highest type of Rorschach response; compare Siipola & Taylor, 1952). (2) Confabulatory, arbitrary and vague responses, or else the most obvious or trivial detail responses (as against only the rare occurrence of responses involving the organization or integration of various aspects of the blot area). (3) Preponderance of personalized or body responses: *Hd, Ad, At, H Obj.*

To adopt Piaget's terminology, the above conceptual responses show that the deaf's thought is still predominantly at the 'transductive stage of development' which 'lies halfway between practical reasoning and truly logical reasoning': the arbitrary responses are examples of 'autistic' thinking, the confabulatory responses of 'intuitive' thinking, and the thought content illustrates its 'ego-centric' nature. That is, the Rorschach test reveals that the deaf children's thinking is insufficiently 'decentred' having not yet been transformed by 'socialization' to the highest level of development, the level of 'logical reciprocity'.

(iii) *Rorschach results in relation to the perceptual organizing processes of the deaf*

The above study of the concept formation aspect of the Rorschach test has revealed a retardation in the deaf's conceptual development. But as Hebb (1949) and others have pointed out, the terms 'percept' and 'concept' refer to processes so intimately related that the term 'conceptual development' can cover both and hence, if the first statement is true, the Rorschach test, being a visual test, should also provide evidence of retardation in their perceptual organizing abilities.

In the previous subsection the Rorschach responses characteristic of the deaf were listed, and it was noted that their formulation involved the minimum of intellectual effort and organization. Those same responses illustrate equally well that their operation was on 'the "least" action perceptual tendency' (Hall, 1951, p. 127) and thus are responses to those structural parts of the ink-blots which involved minimum perceptual organization and integration. Again, it may have been the difficulty they experienced in perceptually articulating the Rorschach blots that caused some of the deaf population to exhibit such behaviour as 'blotting', 'edging' and apparently aimless twisting of the cards. Further, their rejection of the Rorschach cards may have been due not so much to their poverty of associations as to their inability to pierce and/or integrate the blot area.

It appears then, in contrast to the findings of Levine (1948) and Murphy (1952), that the deaf, in their visual perceptual abilities at any rate, are not only not superior to their hearing equals, but may actually be inferior to them. The results therefore are against the old hypothesis of sensory compensation, and thus better perceptual organization; rather, they are evidence for 'a theory of correlation which would hold that abilities correlate positively and where one is weak or missing the others are likely to be depressed' (Pintner, 1941, p. 243).

(iv) *Importance of early environmental experiences*

From the age of five the deaf subjects have been exposed to a scholastic programme with its main emphasis on developing their language and speech. Despite this their language development is far behind that of their 15-year-old hearing counterparts (cf. Bakwin, 1950; Drennan, 1941; Ewing, 1941; Pintner, 1941), and it seems that this lag is permanent. Because of the crucial relationship between language development and complete mental development the cognitive development of the deaf, apparently, will also be retarded. Two main factors could possibly account for this: the present teaching methods and techniques used are inadequate to offset completely the consequences of a deprivation of language experiences in pre-school childhood; and/or irrespective of the standard of formal education, the extreme environmental impoverishment of a wordless pre-school childhood has set rather narrow ultimate limits to their future cognitive development. Experimental evidence (cf. Clarke, Heron, Fetherstonhaugh, Forgays & Hebb, 1951; Forgays & Forgays, 1952; Hebb, 1949, ch. 11; Hymovitch, 1952; Thompson & Heron, 1954) supports this last, although it would seem that the quality of formal education will be the main factor in determining the extent to which the deaf can develop within their already restricted limits (cf. Bindon, 1957). Thus if an early restrictive environment in infancy can have such important and relatively permanent effects in shaping adult behaviour, those concerned with the deaf must direct more attention to the problem of how to present the pre-school deaf child with language experiences at the earliest possible age.

(v) *Wechsler-Bellevue Intelligence Scale Performance Form II and Make-a-Picture Story (MAPS) results for rubella deaf, non-rubella deaf and non-deaf groups*

In addition to the Rorschach test, the Wechsler-Bellevue Intelligence Scale Performance Form II (Wechsler, 1946) and MAPS test (Shneidman, 1948) were administered to all groups.

The results of these tests with the deaf groups, like those of the Rorschach test, revealed that neither cause of deafness, degree of deafness, sex, nor the residential factor, had any specific influence on their performances ($P > 0{\cdot}1$ in the few cases where numerical differences warranted statistical analysis).

On the other hand, the matched rubella deaf group, in comparison with the hearing group, gave on the MAPS test less of Shneidman's 'normal' signs ($\chi^2 = 16{\cdot}86$, $P < 0{\cdot}001$) and more of his 'schizophrenic' signs ($\chi^2 = 6{\cdot}83$, $P < 0{\cdot}01$); with the results in general revealing their social isolation and conceptual and perceptual incongruity. The intelligence test results did not differentiate between these two groups matched for age, sex, educational stream, and approximate socio-economic status as gauged by parental occupation ($P > 0{\cdot}1$). The failure to find a statistically significant difference cannot be regarded as necessarily reliable and may be due to the small N, for the mean I.Q. score and all subtest scores were lower for the rubella deaf group. Their consistently lower subtest scores, taken with their qualitative performances on this particular test, suggest that they may be less effective in comprehending and sizing up a total situation, in differentiating the essential from the inessential details, and in visual analysis and synthesis. (Wechsler's own subtest interpretations adopted here, 1944, ch. 7.)

The results for these two tests therefore support the Rorschach findings in showing that the all-round cognitive development of the deaf has been retarded.

(vi) *Concluding remarks*

The Rorschach results revealed no behaviour difference between rubella deaf and non-rubella deaf children. They did reveal, however, that the outstanding personality characteristics of the deaf (rubella and non-rubella) were their lack of logical reasoning, and their undifferentiated behaviour. This last, it will be remembered, is the basis of that rigidity manifested by immature organisms. Therefore it is suggested that adequate conceptual development may be a necessary pre-condition for the establishment of flexible or differentiated behaviour, which is characteristic of the developed organism. If this is true, any factor that retards cognitive development will also restrict the capacity of the organism to manifest non-rigid and adaptable behaviour. But deafness retards this psychological maturation more than blindness (cf. McAndrew, 1948*a*, *b*). Hence it is not the actual malfunctioning of one of the major senses but the *consequence* of such malfunctioning that appears to be the more crucial factor in restricting full personality development: deafness is more drastic than blindness because it subjects the deaf to a more extreme environmental impoverishment through depriving them of language experiences whose acquisition is the most important factor in the process of adequate socialization—an indispensable condition for complete cognitive development.

REFERENCES

ALTABLE, J. P. (1947). The Rorschach diagnostic as applied to deaf mutes. *Rorschach Res. Exch.* **11**, 74–9.

AMES, L. B., LEARNED, J., METRAUX, R. W. & WALKER, R. N. (1952). *Child Rorschach Responses.* New York: Hoeber, Med. Book Dep. Harper.

BAKWIN, R. (1950). The deaf child. *J. Pediat.* **36**, 668–82.

BECK, S. J. (1931). The Rorschach test in problem children. *Amer. J. Orthopsychiat.* **1**, 501–9.

BECK, S. J. (1937). Psychological processes in Rorschach findings. *J. Abnorm. Soc. Psychol.* **31**, 482–8.

BECK, S. J. (1946). *Rorschach's Test. II—A variety of Personality Pictures.* New York: Grune & Stratton.

BINDON, D. M. (1957). Personality characteristics of rubella deaf children: Implications for teaching of the deaf in general. *Amer. Ann. Deaf.* **102** (in the Press).

BROIDA, D. C. & THOMPSON, G. C. (1954). The relationship between the Rorschach 'insecurity' hypothesis and children's reactions to psychological stress. *J. Personality*, **23**, 167–81.

CLARKE, R. S., HERON, W., FETHERSTONHAUGH, M. L., FORGAYS, D. G. & HEBB, D. O. (1951). Individual differences in dogs: preliminary report on the effects of early experience. *Canad. J. Psychol.* **5**, 150–6.

COX, F. N. & SARASON, S. B. (1954). Test anxiety and Rorschach performance. *J. Abnorm. Soc. Psychol.* **49**, 371–7.

DRENNAN, G. L. (1941). The psychology of deafness in children. *Illinois Med. J.* **79**, 227–32.

EICHLER, R. M. (1951). Experimental stress and alleged Rorschach indices of anxiety. *J. Abnorm. Soc. Psychol.* **46**, 344–55.

EWING, A. W. G. (1941). Defective hearing and mental development (Tests of educational achievement). *Teacher of the Deaf*, **39**, 62–5.

FORD, M. (1946). *The Application of the Rorschach Test to Young Children.* University of Minnesota Press.

FORGAYS, D. G. & FORGAYS, J. W. (1952). The nature of the effect of free-environmental experience in the rat. *J. Comp. Physiol. Psychol.* **45**, 322–8.

GOLDFARB, W. (1944). Effects of early institutional care on adolescent personality: Rorschach data. *Amer. J. Orthopsychiat.* **14**, 441–7.

HALL, K. R. L. (1951). The testing of abstraction, with special reference to impairment in schizophrenia. *Brit. J. Med. Psychol.* **24**, 118–31.

HEBB, D. O. (1949). *The Organisation of Behavior.* New York: Wiley.

HYMOVITCH, B. (1952). The effects of experimental variations on problem solving in the rat. *J. Comp. Physiol Psychol.* **45**, 313–21.

KLOPFER, B. & KELLEY, D. (1946). *The Rorschach Technique.* London: Harrap.

LEVINE, E. S. (1948). *An Investigation into the Personality of Normal Deaf Adolescent Girls.* Ph.D. Thesis, N.Y. Univ. Microfilm of complete manuscript. Univ. Microfilms, Ann. Arbor., Mich. Publ. no. 1156.

MCANDREW, H. (1948*a*). Rigidity and isolation: a study of the deaf and blind. *J. Abnorm. Soc. Psychol.* **43**, 476–94.

MCANDREW, H. (1948*b*). Rigidity in the deaf and blind. *J. Soc. Issues*, **4**, 72–7.

MCNEMAR, Q. (1949). *Psychological Statistics.* New York: Wiley.

MALCOLM, O. Y. (1951). *A Comparative Study of Personality Patterns of Maori and Pakeha Children as Revealed by the Rorschach Technique.* Unpublished M.A. Thesis, C.U.C., University of New Zealand.

MAYMAN, M., SCHAFER, R. & RAPAPORT, D. (1952). Interpretation of the W-B Intelligence Scale in personality appraisal. *An Introduction to Projective Techniques*, ch. 10. Ed. Anderson & Anderson, New York: Prentice-Hall.

MONS, W. (1948). *Principles and Practice of the Rorschach Personality Test.* London: Faber & Faber.

MUNROE, R. L. (1950). The Inspection technique for the Rorschach protocol. *Projective Psychology*, part II. Ed. Abt & Bellak. New York: Knopf.

MURPHY, L. J. (1952). *Assessment of Abilities of Deaf Children.* Ph.D. Thesis, Victoria University of Manchester.

PHILLIPS, L. P. & FRAMO, J. L. (1954). Developmental theory applied to normal and psychopathological perception. *J. Personality*, **22**, 464–74.

PIAGET, J. (1951). *Play, Dreams and Imitation in Childhood.* Part III. London: Heinemann.

PINTNER, R., EISENSON, J. & STANTON, M. (1941). *The Psychology of the Physically Handicapped.* New York: Crofts.

RAPAPORT, D. (1949). *Diagnostic Psychological Testing*, vol. II. Chicago: Year Book Publishers.

RICKERS-OVSIANKINA, M. (1938). The Rorschach test as applied to normal and schizophrenic subjects. *Brit. J. Med. Psychol.* **17**, 227–57.

SALFIELD, D. J. (1950). An attempt at a numerical evaluation of Rorschach test results. *J. Gen. Psychol.* **43**, 305–11.

SHNEIDMAN, E. S. (1948). Manual: *Make a Picture Story Test.* Genetic. Psychol. Monographs, no. 38.

SIIPOLA, E. & TAYLOR, V. (1952). Reactions to ink-blots under free and pressure conditions. *J. Personality*, **21**, 22–47.

SPINLEY, B. M. (1949). *A Study of Two Cultural Groups.* Unpubl. thesis presented for Jacob Joseph Schol., C.U.C., University of New Zealand.

STROOBANT, R. E. (1951). *A Comparative Investigation of the Personality Configurations of Stammerers and Non-Stammerers.* Unpubl. M.A. thesis, C.U.C., University of New Zealand.

SUARÈS, N. D. (1938). Personality development in adolescence. *Rorschach Res. Exch.* **3**, 2–12.

THOMAS, H. F. (1955). The relationship of movement responses on the Rorschach test to the defense mechanism of projection. *J. Abnorm. Soc. Psychol.* **50**, 41–4.

THOMPSON, W. R. & HERON, W. (1954). The effects of restricting early experience on the problem-solving capacity of dogs. *Canad. J. Psychol.* **8**, 17–31.

WECHSLER, D. (1944). *The Measurement of Adult Intelligence.* Baltimore: Williams & Williams.

WECHSLER, D. (1946). *The Wechsler-Bellevue Intelligence Scale. Form II. Manual for Administering and Scoring The Test.* New York: Psychol. Corp.

WERNER, H. (1940). The concept of rigidity: a critical evaluation. *Psychol. Rev.* **53**, 43–52.

WERNER, H. (1946). Abnormal and subnormal rigidity. *J. Abnorm. Soc. Psychol.* **41**, 15–24.

(*Manuscript received* 27 *June* 1956)

PREVIOUS KNOWLEDGE AND MEMORY

By E. C. POULTON

M.R.C Applied Psychology Research Unit, Cambridge

This paper is concerned: (*a*) with the relationship between previous knowledge and memory, (*b*) with differences between recall and recognition, and (*c*) with set for time of memory test while learning. Subjects listened to 144 short statements, half of which were true and half false. They had to rate each statement on a five-point scale for certainty of their truth or falsehood. A test of memory was then carried out for half of the statements, and another test was given a week later for the other half. In each test, half the statements had to be completed (recall), and each of the remaining statements had to be discriminated from a paired statement which resembled it in a standard way (recognition). While rating the statements, half the subjects did not expect a test of memory at all, a quarter were led to expect an immediate test, and the remaining quarter were led to expect a test in a week. After the first test of memory, no subjects were led to expect a second test. A control group of subjects took a test of 'memory' without hearing the statements.

It was found that a subject was more likely to be certain that a false statement was false, than that a true statement was true. For both true and false statements, memory was best for statements about whose truth the subject was certain, and worst for statements about whose truth he had no idea. The control subject was much less successful than the experimental subject, but tended to score on the same statements. Previous knowledge can thus assist both genuine memory and also guess-work.

With memory for a particular word as a criterion, there was no difference between recall and two-choice recognition when the word was in a statement about whose truth the subject was certain and correct, and memory was tested almost immediately. But recognition was more successful than recall when the word was in a statement about whose truth the subject had no idea, or when memory was tested after a week, although differences were only significant in the latter case.

Recognition was found to be a less sensitive measure of previous knowledge than recall. Success in recognition was determined to a considerable extent by factors intrinsic to the test of recognition. Part of the reduced sensitivity of recognition was attributed to this cause.

Immediate memory was best when a test of memory was expected immediately, and worst when no test was expected, but the differences were not significant. If anything, the trend for memory after a week was in the reverse direction.

I. INTRODUCTION

The primary aim was to determine the size of the effect upon memory, for meaningful statements, of previous knowledge in the field of the statement. The need for this investigation is virtually self-evident for, in a sense, a statement can only have meaning to a subject if he has some relevant previous knowledge; and differences between the learning of meaningful and meaningless materials are well established (McGeoch, 1942). Yet the need for an adequate control of previous knowledge in the design of learning experiments with prose has not always been realized. Effects due to differences in previous knowledge may thus have been ascribed to other causes (Edwards, 1941; Levine & Murphy, 1943; Vernon, 1950; Belson, 1952).

There have been numerous studies relating previous knowledge to success in learning, but most of them have followed a standard pattern: use has been made of nonsense syllables, or single words paired together; and either an artificial amount of previous familiarity has been introduced at the start of the experiment, or familiarity has been inferred from available group norms, such as those on the association-value of nonsense syllables. In general, prose has been avoided; and when prose has been used, no independent measure has been obtained of each particular subject's everyday experience with each particular item of information contained in the prose.

One subsidiary aim of the present experiment was to investigate two comparisons between tests of recognition and tests of recall. (*a*) When the same criterion of memory is used for both tests, is there a difference between recall and two-choice recognition? Also, what effect do difficult experimental conditions, such as unfamiliar material and prolonged periods of retention, have upon the size of this difference? No experimental comparison using the *same* criterion of memory for recognition and recall yet appears to have been published. (*b*) Tests of recognition appear to be less sensitive to changes in the independent variables of the learning situation than do tests of recall (Postman, Adams & Phillips, 1955). Why is this so, and is it possible to improve the sensitivity of tests of recognition?

A second subsidiary aim was to investigate the size of the influence *on learning* of set for time of memory test. Previous experiments suggest that memory is best when tested at the expected time (McGeoch, 1942). But the results only appear to be valid when the expected time is immediately after learning. For when, after learning, the subject continues to expect to be tested at some specific time in the future, rehearsal during the interval has clearly not been eliminated as a possible cause of the differences found. The present experiment eliminated this ambiguity.

A group of incidental learners was included for the sake of comparison. Incidental learners tend to do less well on memory tests than intentional learners, but the differences are only marked when the learning task and/or the memory test are difficult (Postman *et al.* 1955).

II. Method

Four lists, each of thirty-six statements, were recorded on magnetic tape. A statement was read every 10 sec. One of the most familiar statements was: 'The stern may be the rear of a steamer'. The least familiar was: '*Sinanthropus* was dug up in Australia'. One of the statements which obtained the most qualified ('probable') ratings was: 'A tablespoonful is four teaspoonfuls'. About 30% of the statements could be classified as scientific, 20% as historical, and 15% as geographical. They were selected to be sufficiently different from one another in meaning to minimize the number of confusions between them.

The statements ranged in length from 3 to 9 words, with an average of 6·1 words. Each started with a title, e.g. the stern, and contained two key words, e.g. rear and steamer. About 20% of the titles contained a word which was probably unfamiliar to half or more of the subjects (see Results). No two key words were the same. About 80% of them were included in the commonest 2000 words of the English language (Thorndike & Lorge, 1944). (Slight modifications were made in cases where English and American usage differ, e.g. petrol and gasolene.) The remaining 20% were included in the commonest 7000 words. Only one such less common word was included in any one of the statements. These statements were divided equally between the four lists, and were distributed randomly within lists.

About 8% of the statements were phrased in the negative. They were distributed randomly between and within lists. Half the statements in each list would generally be considered to be true, the other half false. About 95% of the true statements were simple statements of fact, whereas only about 40% of the false statements were either simple statements of fact containing a negative, or in some sense the opposite of the truth. The remaining false statements were not related to the truth in any simple or uniform way. The order of true and false statements within lists was randomized.

(i) *Self ratings of previous knowledge.* The experiment was carried out in three parts. The first part was presented as a new kind of general knowledge test. The statements were played over loudspeakers. The subject had a list of the titles of the statements in front of him. After each title there were five boxes. For half the subjects they were labelled in order: certainly true, probably true, no idea, probably false, certainly false. For the other half of the subjects the boxes were labelled in the reverse order. After hearing each statement, the subject had to put a tick in the relevant box. The instructions were phrased to prevent him from being either over-cautious or over-rash. The difficulty of the statements was so graded that about one-third of them should have fallen into each of the three categories: certain, probable, and no idea, and the subject was told this. (Difficulty was estimated in a pilot experiment with twenty-four other subjects.) If the subject did not hear a statement, he was instructed to leave a blank, not to place a tick in the box labelled no idea.

The experimenter started by reading out six examples. In each case, he stated how he would respond, and why. This was followed by a practice of twelve statements, played over the loudspeakers. The subject's responses were subsequently checked and discussed by the subject and experimenter together. If there was to be a warning that a test of memory would follow the test of general knowledge (see below), it was then given. There was a break of about one minute after the first two lists of statements had been heard.

(ii) *Immediate test of memory*. Immediately after the test of general knowledge, there was a test of memory for the first two of the four lists of statements. To twenty-four subjects, the incidental learners, this came as a complete surprise. Twelve subjects had been warned to expect an immediate test of memory, and the remaining twelve had been warned to expect a test in a week's time, but not immediately. In no case was the nature of the test of memory disclosed in advance.

The titles of one list of statements were presented on an otherwise blank sheet, and the subject had to complete as many of the statements as he could. He was told to give the original wording if possible, and at least to give the correct meaning. The statements of the other list were then presented in full. Each statement had a similar statement typed above or below it, from which it had to be distinguished. Half the originals were typed first, and half second, the order of pairs being randomized within lists. The subject was made to guess if he was not sure. But he was told that he should be able to decide in almost all cases without guessing. Half the subjects performed the recall test first, half the recognition test. Unlimited time was available for both tests.

In the test of recognition, the alternative version of a statement had the same title as the original. Half the alternative versions had one different key word, and the other half had both key words different. Each of these subgroups was divided into three, according to meaning. The meaning of the alternative was either more or less the same as the original, in some sense the opposite, or as far as possible unrelated. The alternative statements were carefully paired with the originals. Thus a statement which contained a less-common key word was paired with a similar statement. In addition, the alternative statements were selected to make alternative lists, which were similar to the original lists in all the respects required by the experimental design. Thus the alternative lists could also have been used for the test of general knowledge, and this was originally the intention.

(iii) *Delayed test of memory*. This was carried out one week later, and was similar to the immediate test. It used the last two of the four lists of statements heard. This test came as a complete surprise to all subjects, and thus minimized rehearsal. The subjects who had been warned to expect a test in a week's time, had been told at the time of the immediate test that the immediate test was actually to be their test.

(iv) *Order of lists*. Twelve of the twenty-four incidental learners heard the four lists of statements in the order 3 4 1 2. The remaining thirty-six subjects all heard the lists in the order 1 2 3 4. The lists were constructed to be of approximately equal difficulty. Of each subgroup of twelve subjects, six had to recall lists 1 and 3 and recognize lists 2 and 4. The other six had to recognize lists 1 and 3 and recall lists 2 and 4. Memory for list 1 was always tested before memory for list 2, and memory for list 3 before list 4. (This design confounds the difference between immediate and delayed memory with the order of hearing the lists.)

(v) *Control group*. A group of twenty-four subjects had to take recall and recognition tests without having heard the statements. In a preliminary training period each subject copied down the six examples, and the nature of the statements was explained to him. Afterwards he had to complete ('recall') the twelve practice statements. His attempts were examined by the experimenter and himself, and the actual statements used in the practice were then read to him. In the test which followed he had first to complete the statements of one list, and then to select ('recognize') one of each pair of statements of one other list. He was told to judge between the pairs of statements on their merits, and not just to guess at random. Conditions were otherwise the same as for the memory tests of the experimental groups. Twelve of the subjects worked with lists 1 and 2, the remainder with lists 3 and 4. Of each subgroup, half completed one list and the other half completed the other list.

(vi) *Subjects.* The seventy-two subjects were all naval ratings, aged between 17 and 27 years, with a median age of 20. Their scores on intelligence test AH4 (Heim, 1955) ranged from 43 to 116 with a median of 72. Most had left school at 15, and had been apprentices or had taken miscellaneous jobs before joining the Navy. They were tested in groups of six. The experimental groups were of approximately similar composition.

(vii) *Scoring.* Recall was normally scored by two methods (see Table 2 or Table 3). (*a*) In one method a statement was considered to be correctly recalled if the two key words were given more or less correctly. The remaining words, and the meaning of the recalled statement, were neglected in this method. The exact form of the key words (e.g. adjective, noun, or verb, singular or plural, present or past tense) was also neglected, since this depended on the structure of the sentence. Only about 60% as many statements would have been scored as correctly recalled by this method if the exact wording of the original had been required. (*b*) In the second method, a statement was considered to be correctly recalled if the meaning was more or less correct. The recalled version had to reflect the meaning of one of the two key words, and not to alter too grossly the general sense of the other key word. Excess information was not penalized either if it did not alter too grossly the meaning of the original statement, or if the meaning of the original was ridiculous anyway. In order to satisfy the experimental design, it was necessary to ensure that the criteria used for the various statements were equally strict. However, a certain degree of apparent arbitrariness cannot be avoided in matching differences in meaning. If the recalled version had had to reflect the meaning of both key words, only about 80% as many statements would have been scored as correct by this method.

In comparing recall with recognition, the criterion used was remembering one particular key word. This was the key word which was changed in the set of alternative versions with more or less the same meaning as the original statement, and with one identical key word. In the test of recall, the criterion for this key word was the same as that described in (*a*) above. (The key word had only to be recalled sufficiently correctly for the experimenter to be reasonably certain which one of the two alternative words used in the test of recognition was intended.) This criterion of memory restricted the comparison to one-sixth of the statements.

(It is not possible to compare recall and recognition on memory for meaning alone, without altering in the test of recognition the wording of both the original statement and the alternative version, for an alternative version with a different meaning necessitates some change in wording. And a change in wording as well as in meaning favours recognition, unless the 'original' statement has also undergone a corresponding change in wording—see Discussion.)

(viii) *Calculations.* An index of general knowledge was computed by subtracting the certain and probable true-false ratings which would generally be considered to be incorrect, from those that would generally be considered to be correct. A recognition score was also computed, by doubling the number of errors in the test of recognition, and subtracting this product from the greatest possible score. (This recognition score is the closest approximation which can be obtained to the amount of genuine memory shown in the test of recognition. The underlying assumptions are (*a*) that errors are wrong guesses, and (*b*) that corresponding to every wrong guess there was on average a guess which was right, since there were only two possible choices and the subject was forced to select one of them. On this view, doubling the errors eliminates on average the correct guesses.)

The significance of the difference of each pair of means was assessed separately (*a*) against the variability between subjects, and (*b*) against the variability between statements. The level of the less significant of the two tests was taken. This test was almost always (*b*). Two-tailed tests were used except where stated. As compared with more conventional methods, this method tends to underestimate significance. For Tables 2 and 5 the difference between paired proportions was assessed, since the incidence of the self ratings of previous knowledge varied as well as the memory score. A simple sign test was used for each pair of proportions, and the numbers of positive and negative outcomes were compared by χ^2.

The means of the experimental subjects were computed from the results of all forty-eight subjects. However, the levels of significance were generally computed only from the results of the thirty-six subjects who heard the lists in the order 1 2 3 4. Exceptions to this rule are the significance levels in Table 1, which are for all forty-eight subjects, and others specifically mentioned in the text. (The memory scores of the twelve subjects who heard the lists in the order 3 4 1 2 could not be combined with the remainder when carrying out test (*b*).)

III. RESULTS

(i) *Previous knowledge.* The distributions of the self ratings of previous knowledge are shown in Table 1. The subject was more likely to be certain that a statement which would generally be considered to be false was false, than that a true statement was true. The order of the five boxes (see Method) made no difference to the ratings. About 20 % of the statements received 'no idea' ratings from half or more of the subjects. Practically all these statements contained an unfamiliar word in the title. They accounted for about 55 % of all the 'no idea' ratings. The extreme values of the index of general knowledge for different subjects were 20 and 68 %. For brevity, the data on individual differences will not normally be given, but the scatter was always large.

Table 1. *Truth of statement and self rating of previous knowledge*

	Percentage of statements rated				
	Certain		Probable		
	Correct	Incorrect	Correct	Incorrect	No idea
True statements	31	5	27	12	25
False statements	39	3	20	11	27
Difference	−8	2	7*	1	−2

* This difference is significant at the 0·001 level of confidence. It is significantly different from the difference for certain correct at the 0·01 level.

(ii) *Memory.* Table 2 gives the memory scores for the different self ratings. The statements about which the subject was certain were remembered best, and the statements about which he had no idea were remembered least well. This trend was shown by both the true and the false statements. In addition, the statements which received the conventional true-false rating ('Correct' in Table 2) were remembered better than the statements which received the corresponding unconventional rating ('Incorrect').

Table 2. *Self rating of previous knowledge and memory*

	Self rating of previous knowledge				
	Certain		Probable		
Criterion of memory	Correct (%)	Incorrect (%)	Correct (%)	Incorrect (%)	No idea (%)
Memory after 30 min.					
Two key words reproduced*	56	41	44	33	16
Meaning more or less reproduced†	89	76	76	65	33
Recognition score‡	92	84	90	94	79
Memory after 1 week					
Two key words reproduced§	18	16	13	11	2·2
Meaning more or less reproduced¶	45	30	31	26	9
Recognition score**	76	62	70	66	40

* Combined correct and incorrect certain–probable, $P = 0{\cdot}05$. Probable correct different from probable incorrect at 0·05 level, and from no idea at 0·01 level.

† Probable correct different from both certain correct and no idea at 0·01 level.

‡ Probable incorrect–no idea, $P = 0{\cdot}01$.

§ Certain correct–no idea, $P = 0{\cdot}01$.

¶ Certain correct–probable correct, $P = 0{\cdot}01$.

** Certain correct–no idea, $P = 0{\cdot}05$.

The meanings of the true statements were rather better recalled after 30 min. than the meanings of the false statements ($P=0{\cdot}02$). There was no difference between memory for the different kinds of false statements. Only about 5% of the recalled versions of the false statements had a meaning in some sense opposite to that of the original. The corresponding figure for the true statements was 3%. Mishearings in the original test of previous knowledge appeared to account for less than 1% of the errors in recall.

Memory after 30 min., for the 20% of statements which contained an unfamiliar word in the title, was about the same as the results in Table 2 for the statements rated 'no idea'. After a week memory for the statements with an unfamiliar word in the title was rather worse in every case than the results in Table 2 for the 'no idea' statements.

In the test of recall after 30 min. none of the forty-eight subjects obtained full marks by either criterion. Only 3% of the statements to be recalled by eighteen subjects always had the two key words reproduced correctly, and only 18% of the statements always had the meaning more or less correctly reproduced. In contrast, in the test of recognition after 30 min., 12% of the subjects, and 65% of the statements to be recognized by eighteen subjects, produced no errors. After a week 17% of the statements were still always recognized correctly, although none of the subjects obtained full marks for recognition.

(iii) *Set for test of memory.* By all three criteria of memory, immediate memory was best when the immediate test of memory was expected, and worst when no test was expected. Expecting a test in a week gave intermediate results. If anything, the trend for memory after a week was in the reverse direction, although the differences were smaller. However, none of the differences due to set for test of memory was significant with the stringent tests of significance used ($P>0{\cdot}05$).

(iv) *Guess-work.* Table 3 shows the memory scores after a week for all statements combined, together with the scores for the guesses of the twenty-four control subjects, who did not hear the statements. Also given are the τ correlations by statements between the scores of the control subjects, and the scores after a week of the twenty-four incidental learners. The memory scores after a week were significantly better than guess-work, although all subjects tended to score on the same statements in the tests of statement completion.

Table 3. *Memory after* 1 *week compared with guess-work*

Criterion of memory	Memory after 1 week (%)	Guess-work (%)	τ by statements, memory after 1 week (incidental learners)–guess-work
Two key words reproduced	12	1*†	0·30‡
Meaning more or less reproduced	29	9*§	0·37¶
Recognition score	64	4·5*	0·09

* Memory after 1 week (incidental learners only)–guess-work, $P=0{\cdot}001$.
† Guess-work different from zero at the 0·05 level.
‡ $P=0{\cdot}02$.
§ Guess-work different from zero at the 0·001 level.
¶ $P=0{\cdot}001$.

For the control subjects only 6% of the completions of the false statements had a meaning in some sense opposite to that of the original. When the two alternatives in the test of statement selection had the same meaning, the control subjects selected the correct alternative 25% more often than would be predicted by chance ($P=0{\cdot}001$).

(v) *Recognition.* The effect on recognition score (see Calculations) of the relationship between the original and the alternative statement is shown in Table 4. Over 40% of the errors in recognition after 30 min. were attributable to 8% of the statements. These statements were typed below the alternative statement, contained one of the same two key words, and had more or less the same meaning as the alternative.

Table 4. *Effect of recognition test on recognition score*

Relationship between original and alternative statement	Recognition score after			
	30 min. (%)	*P*	1 week (%)	*P*
Original typed: above	96	0·05	70	0·05
below	81		54	
Identical key words: 1	82	0·05	52	0·001
0	93		74	
Meaning: same	79	0·05	80	
opposite	90		78	>0·1
different	94		68	

In both the immediate and the delayed tests of recognition, the statements paired with the different kinds of alternative all showed the same trend with previous knowledge. The nature of the alternative only influenced the mean recognition score, and the size of the trend. In the test of recognition after a week, the statements paired to alternatives with only one key word changed were the most sensitive statistically to differences in previous knowledge. For these statements the difference between certain correct and no idea was significant at the 0·02 level (cf. Table 2).

Table 5 compares two-choice recognition with recall, when the same criterion of memory was used in both tests. Recognition was either as successful as recall, or more successful.

Table 5. *Recall compared with recognition*

Criterion of memory	Self rating of previous knowledge			
	Certain correct* (%)	Probable correct (%)	No idea* (%)	All ratings combined† (%)
Memory after 30 min.				
One specific key word reproduced	70	44	22	50
Recognition score when alternative statement had only this key word changed	69	78	34	64
Memory after 1 week				
One specific key word reproduced	33	19‡	3·6	22
Recognition score when alternative statement had only this key word changed	72	72	21§	55

* After 1 week recall–recognition, $P = 0{\cdot}05$, with a one-tailed test.
† After 1 week recall–recognition, $P = 0{\cdot}01$.
‡ Certain correct–probable correct, $P = 0{\cdot}05$, with a one-tailed test.
§ Certain correct–no idea, $P = 0{\cdot}05$, with a one-tailed test.

(vi) *Experimental material.* There was no difference in any respect between the statements containing the less common key words and the remaining statements. The four lists of statements gave similar mean ratings for familiarity. In addition, the memory scores of the two groups of incidental learners showed that the two halves of the material were about equally difficult to remember.

(vii) *Incidental learning and intelligence.* The τ correlation of the total score on intelligence test AH4 (Heim, 1955) with the index of general knowledge was 0·55 for the subjects who expected a test of memory, and 0·36 for the subjects who did not expect one ($P = 0{\cdot}01$ or better). In order to determine the direct correlation of intelligence with memory, the correlations of the index of general knowledge with intelligence on the one hand and with memory on the other had to be partialled out. When this was done for the incidental learners, the total on AH4 correlated 0·30 with immediate memory, and 0·37 with delayed memory. The corresponding values for the subjects who expected a test of memory were 0·01 and 0·04. Recall of the two key words was used as the measure of memory. Thus with subjects who expected a test of memory, intelligence was reflected principally in the index of general knowledge. Whereas with the incidental learners, intelligence was about equally reflected in the index of general knowledge and in memory. These results are in line with the view that intelligence is closely related to incidental learning, and thus to general knowledge, but is not so closely related to the ability to learn by intention (Carr & Kingsbury, 1938, p. 373).

IV. Discussion

(i) *Previous knowledge assisted memory.* Table 2 shows that, when the subject was certain and correct about the truth of a statement, he was more likely to recall its meaning than when he was less certain of its truth ('Probable Correct' in Table 2). This difference was significant at the 0·01 level both when memory was tested after about half an hour, and when it was tested after a week. Similar trends were found when only the presence or absence of the two simple key words in each reproduction was considered, and in the test of recognition, although these trends were not significant. The differences are interpreted as follows: Degree of certainty about the truth of a statement was related to the amount of previous knowledge in the field of the statement, and amount of previous knowledge was related to success in the tests of memory.

Table 2 also shows that statements about whose truth the subject was certain and correct, were recalled after 30 min. 2·7 or 3·5 times better than statements about whose truth he had no idea, depending on the criterion used. After a week the ratios were 8·2 and 5·0. On this extreme comparison both criteria of recall showed significant differences at both time intervals, and recognition showed a significant difference after a week. The memory scores for the 20% of statements which contained an unfamiliar word in the title were much the same as the scores in Table 2 for the statements rated 'no idea'. In fact, about 55% of all the 'no idea' ratings went to these statements. Hence the difference in memory between the statements rated 'certain' and 'no idea' is to some extent analogous to the difference found between an old 'stimulus' and a new 'stimulus' in experiments on the learning of nonsense syllables by the method of paired associates.

It was also found that statements which would generally be considered to be true were rather better recalled (though not necessarily better recognized) than false statements, although the subject was more likely to be certain in his rating of false statements (Table 1). Most knowledge is probably stored in a positive rather than a negative form. Thus the false statements were presumably less directly related to previous knowledge than were the true statements. (The difference was not due simply to recalling true meanings which were the opposite of the correct false meanings. For the proportion of opposite meanings recalled was about the same for the true and false statements.)

(ii) *Previous knowledge assisted statement completion by guess-work.* Table 3 shows that the control subjects, who had to do the tests of 'memory' by guess-work, scored significantly better than zero on statement completion ('recall'), although not on statement selection ('recognition'). In addition, they tended to score on the same statements as the experimental subjects. This was to have been expected, since the guessing of the control subjects in completing the statements was presumably directed by previous knowledge.

Previous knowledge may also have directed the guessing of the experimental subjects, as well as assisting them in learning and recall. Table 3 suggests that guess-work could not have been much help in spotting the actual wording of the statements to be recalled. This was because many statements did not contain the customary key words, owing to the requirement that the same key word must not appear in two statements. But guess-work may have helped considerably with the meanings of the statements, especially in the test after a week. In the present experiment the information in different statements was more or less independent. In normal prose guessing may be a more important factor in successful 'memory'. For in prose there is generally no limitation on the incidence of key words, and consecutive meanings are not independent of each other.

A small but significant 'recognition' score was obtained by guess-work when the two alternative statements had the same meaning. The correct alternatives tended to have the more common combinations of words. This small bias in the experimental design should not have occurred.

(iii) *Relevance to some previous memory experiments.* Edwards (1941) read a controversial passage to two groups of students with widely different political views. In a subsequent test of recognition, he found that both groups tended to remember the points which corresponded to their political outlook better than the points which conflicted with it. Following an hypothesis derived from Sherif (1936), he concluded that political frames of reference acted selectively on the learning of, and memory for, the controversial material. But in his design differences in previous knowledge between the two groups were not balanced. There was an additional finding that the over-all memory score of the group containing the more politically experienced students was higher than the score of the group containing the less politically experienced students. This strongly suggests that previous knowledge did influence the results. Differences in previous knowledge congruent with the political frames of reference cannot therefore be excluded as the principal determinant of the differences found. A similar criticism applies to the experiment of Levine & Murphy (1943).

The present results also suggest that an uncontrolled test of memory is not an appropriate measure of the comprehensibility of a passage. This measure has recently been used on educational talks for adults broadcast by the B.B.C. (Vernon, 1950; Belson, 1952). But this measure does not distinguish between what was known before hearing the passage, and new information which has been acquired by listening. Comprehension is confounded with previous knowledge. Some confounding is inevitable, for memory is, as it were, loaded *against* new knowledge (Table 2). But the use of a control group of subjects similar to the experimental group, which takes the same test without having listened to the passage (cf. Table 3), will at least indicate what could have been achieved by previous knowledge alone.

(iv) *Truth and certainty.* Table 1 shows that the subject was more likely to be certain in his rating of a statement which would generally be considered to be false, than in his rating of a true statement. His greater caution with the true statements was probably not due to having available less relevant previous knowledge. For the proportions of incorrectly rated statements, and the proportions of statements about which he had no idea, were practically the same for the two types. The true statements must therefore have seemed more ambiguous to the subject than the false statements. (About half of both types of statement appeared to the experimenter to be ambiguous in some degree. This was because they were selected to be short and simple, and were thus oversimplifications. Brevity and simplicity without ambiguity is the target of all designers of so-called objective tests (Ruch, 1929; Vernon, 1940).)

(v) *Variables which affected the relationship between success in recognition and success in recall.* In comparing the relative difficulty of tests of recall and recognition, the same criterion of memory must be used for the two tests. The criterion used in Table 5 was remembering one particular key word (see Scoring). The justification for using recognition score to represent the amount of genuine memory shown in the test of recognition has been outlined above (see Calculations). With statements about whose truth the subject was certain and correct, recall and two-choice recognition were about equally successful in the test after 30 min. But as conditions became more difficult, with less familiar statements and/or delayed memory, two-choice recognition was the more successful.

No previous experiment comparing recall with recognition appears to have used the same criterion of memory for the two tests. Differences in criterion inevitably favour recognition. For the easiest test of recall, the reproduction of a single item more or less correctly, corresponds to the most difficult test of recognition, the selection of the correct version when the alternative(s) have only this single item different. The more that is demanded concurrently in a test of recall, the *more difficult* it becomes. Whereas the more that is changed concurrently in a test of recognition, the *easier* it becomes (see Table 4).

As an illustration, we can consider an experiment by Postman, Jenkins & Postman (1948), who claimed to have demonstrated that recognition was easier than recall. For recall they demanded perfect reproduction of three-letter nonsense syllables, and for recognition they presented four choices like the following: QUN, MUN, JIC and JIV. But, as compared with recall, recognition was too easy, for two letters (UN, JI) were varied concurrently. The criterion of memory which corresponds to the perfect recall of all three letters, demands a recognition test with at least the following eight choices: ABC, AB3, A2C, 1BC, A23, 1B3, 12C and 123.

(A test of recognition can vary along two distinct dimensions. First, it can test memory for more than one item. And secondly, for each item to be tested, there can be more than one alternative. If memory for each item is to be tested independently (as is necessary when making the correct comparison between recognition and recall), and every alternative is to vary independently of every other alternative, the total number of alternatives will be $(r+1)^n$, where r is the number of alternatives for each item, and n is the number of items for which memory is to be tested. Thus in the modified Postman, Jenkins & Postman experiment, with one alternative for each of three items, the total would be $2^3=8$. Recognition becomes more difficult with more alternatives, whether the number is increased to test memory for each item independently, or simply to increase the total number of choices for each item (Postman, 1950).)

(vi) *Why recognition was a less sensitive measure than recall.* In Table 2 recognition after 30 min. showed only one rather atypical significant difference with previous knowledge, whereas both the measures of recall showed the predicted significant differences. For short-term memory, recognition was thus a less sensitive measure of previous knowledge than was recall. A somewhat similar finding has been reported by Postman *et al.* (1955) for nonsense syllables. One reason for the relatively low sensitivity of recognition may have been that the test was too easy after 30 min., for 12% of subjects and 65% of statements escaped without any errors. Table 2 shows that after a week recognition did just give the over-all significant difference ($P=0{\cdot}05$) with previous knowledge which was predicted. This difference became more significant ($P=0{\cdot}02$) when only the more difficult pairs of statements containing one identical key word were considered.

Table 4 suggests a second reason for the relatively low sensitivity of recognition, which applies to both short-term and longer-term memory. Success in recognition depended to a considerable extent upon two factors which were specific to the test of recognition, in addition to the experimental variables. These factors were the order of presentation of the old and new items, and the similarity of the new items to the old. If a test of recognition is to be made as sensitive as possible, either the influence of these factors must be held constant, or the part of the variance resulting from them must be isolated. (In the present experiment the number of statements in each subgroup was too small to give any significant differences with previous knowledge after 30 min., and the form of the results ruled out analysis of variance.)

Table 3 indicates a contributory cause of the difference in sensitivity between recognition, and recall of the meaning of the statements. Guessing may have somewhat increased the 'recall' score of the meanings of the more familiar statements, especially in the test after a week. However, Table 3 suggests that guessing did not appreciably affect the recall score of the two key words in each statement.

The influence of Sir Frederic Bartlett, F.R.S., is shown clearly in this paper. Prof. T. A. Ryan pointed out the need for an adequately designed experiment to investigate the influence of set for time of memory test upon learning. Prof. P. E. Vernon loaned me his copy of his 1950 report. Mr M. Stone gave advice on statistical methods. The subjects were supplied by the Royal Navy. Financial support from the Medical Research Council is also gratefully acknowledged.

REFERENCES

BELSON, W. A. (1952). An inquiry into the comprehensibility of 'Topic for Tonight'. B.B.C. Listener Research Report, no. 1080.

CARR, H. A. & KINGSBURY, F. A. (1938). The concept of ability. *Psychol. Rev.* **45**, 354–76.

EDWARDS, A. L. (1941). Political frames of reference as a factor influencing recognition. *J. Abnorm. Soc. Psychol.* **36**, 34–50.

HEIM, A. W. (1955). *Manual for the Group Test of General Intelligence AH4.* London: National Foundation for Educational Research.

LEVINE, J. M. & MURPHY, G. (1943). The learning and forgetting of controversial material. *J. Abnorm. Soc. Psychol.* **38**, 507–17.

MCGEOCH, J. A. (1942). *The Psychology of Human Learning.* New York: Longmans, Green.

POSTMAN, L. (1950). Choice behavior and the process of recognition. *Amer. J. Psychol.* **63**, 576–83.

POSTMAN, L., ADAMS, P. A. & PHILLIPS, L. W. (1955). Studies in incidental learning. II. The effects of association value and of the method of testing. *J. Exp. Psychol.* **49**, 1–10.

Postman, L., Jenkins, W. O. & Postman, D. L. (1948). An experimental comparison of recall and recognition. *Amer. J. Psychol.* **61**, 511–19.

Ruch, G. M. (1929). *The Objective or New-Type Examination.* Chicago: Scott, Foresman.

Sherif, M. (1936). *The Psychology of Social Norms.* New York: Harper.

Thorndike, E. L. & Lorge, I. (1944). *The Teacher's Word Book of 30,000 Words.* New York: Teacher's College, Columbia University.

Vernon, P. E. (1940). *The Measurement of Abilities.* London: University of London Press.

Vernon, P. E. (1950). The intelligibility of broadcast talks. B.B.C. Listener Research Report, no. 2328.

(*Manuscript received* 26 *May* 1956)

MEASURES OF SUBJECTIVE PROBABILITY

ESTIMATES OF SUCCESS IN PERFORMANCE IN RELATION TO SIZE OF TASK

BY JOHN COHEN, E. J. DEARNALEY AND C. E. M. HANSEL
Department of Psychology, University of Manchester

The question, which arose in an earlier investigation, of the transposability of estimates made on different ranges ('out of five', 'out of ten', etc.) is subjected to experiment. The degree of transposability is found to be a function of age.

I. INTRODUCTION

By subjective probability we understand a measure of mental states of uncertainty which extend from complete uncertainty at one extreme to complete certainty at the other. These states are implicit in all human behaviour, in every perception, thought, judgement, feeling and action, and they assume particular significance in all manner of risk-taking, including gambling, in judgements about the outcome of future events, and in our orientation to the future as partly expressed in our expectations, hopes and fears. One of the aims of the study of subjective probability is to make these implicit states of mind explicit, so that they can be measured.

In our recent studies of risk-taking (1956*b*) we have measured subjective probability (ψ) by asking each person to guess how many times he would succeed in a task out of five attempts. The number five was used because each subject was actually to make five attempts, so that his estimates could be compared with his performance. In this particular situation it would have been tedious for the subjects to make more than five attempts at each level of difficulty. In other situations, subjects have been asked how many times out of ten or out of one hundred they thought they would succeed. Measures of subjective probability were obtained by taking the ratio of the number of expected successes to attempts.

This procedure raises the question whether different values for subjective probability would be obtained if the subject had to express his successes in relation to a larger or smaller number of attempts. An experiment was therefore designed to answer this question. Children were asked to estimate the number of times they would succeed in throwing a ball to hit a target when the number of attempts was set at each of the following values: 1, 5, 10, 100, or 1000. This enabled us to see whether the ratio of estimated successes to attempts would be constant regardless of the number of attempts. These particular values were chosen because they were within the range of the experience of the school-children to be studied. For example, the range 0 to 10 is associated with football, and 0 to 500 with the game of cricket. Children in English schools are generally well informed on matters relating to these two national sports. The targets were varied in size so as to yield eight levels of difficulty and ensure a diversity of values of subjective probability on each range.

II. PROCEDURE

Four classes of school-children consisting of 62 children aged 9–10 years, and 68 children aged 13–14 years acted as subjects. They had to imagine they were throwing a table-tennis ball to hit a target through a circular aperture 4 ft. away. The aperture and target were placed on an easel in front of a class and one child was asked to make a trial throw in order to illustrate the hypothetical task. This trial throw was made when the aperture was 2 ft. 3 in. in diameter. The attempts always succeeded. A series of apertures was presented with diameters of eight different magnitudes, viz. 2·25, 1·75, 1·50, 1·25, 1·00, 0·75, 0·50, and 0·25 ft. Each child received a booklet of six pages. On each page the child recorded his estimates of his expected successes in aiming at each of the apertures in a given number of attempts (1, 5, 10, 100 or 1000), a different range being set on each page. The questions, printed at the top of the page, were of the following form: 'You have a thousand shots at the target. How many times do you think you will hit it?' Each class of children was divided into five subgroups, each subgroup receiving the tasks in a different order. The order in which the five ranges were presented was different for each of the five subgroups as set out in Table 1.

Table 1. *Order of presentation of ranges*

Subgroups of children				
I	II	III	IV	V
0–1	0–5	0–10	0–100	0–1000
0–1000	0–100	0–5	0–10	0–1
0–100	0–10	0–1	0–1000	0–5
0–5	0–1000	0–100	0–1	0–10
0–10	0–1	0–1000	0–5	0–100

The apertures were presented in random order with respect to size. Two of these apertures (0·75 and 1·50 ft.) were presented twice to provide a measure of the reliability of the estimates. Each child thus made fifty estimates.

At the foot of each page, after ten estimates had been made, the subjects answered the following question: 'How many of the ten estimates you have made do you think would be correct if you *actually* carried out the task each time and not just *imagined* that you did it?' Finally, on the sixth page, the children answered the following questions given and explained verbally: (i) Which of the five kinds of question did you find easiest to answer?; (ii) Which did you find the hardest to answer?; (iii) Which of the five questions (or ranges) would you use if you were explaining to a friend how hard it is to throw the ball at this target through the aperture in front of you? (The aperture was set at a diameter of one foot.)

The five different ranges were written on the blackboard to make it easier for the children to reply to the above questions

III. RESULTS

The median estimates of success made by the two age-groups are shown, expressed as ψ-values, in Figs. 1 and 2. Medians rather than means are shown since, in some cases, children made obvious errors because they sometimes carried over one range to the next page. Rather than reject their estimates entirely, and thus introduce a factor of selection, we have employed the median which reduces the effects of large isolated errors on the results. The first point that emerges from Figs. 1 and 2 is that subjective probabilities, as measured by estimates of success, vary systematically with the number of hypothetical attempts. The effect is similar in both age-groups but smaller in the older group. Estimates of success decline as the number of hypothetical attempts increases.

The second point is that subjective probabilities of success vary with the difficulty of the task, which is presumably related to the size of the target. In the younger group, estimates of success decline progressively in all ranges as the task becomes harder. It may be that the range of 'out of a thousand' is beyond the competence or experience of some of these 9-year olds. In the older group the estimates also decline, but not to the same extent; they seem more stable and less affected by the range on which they are expressed. The extent to which the ranges 'out of five' and 'out of a hundred' are equivalent and therefore transposable is brought out in Table 2. There is a marked tendency in the older group to move towards the 'expected' values.

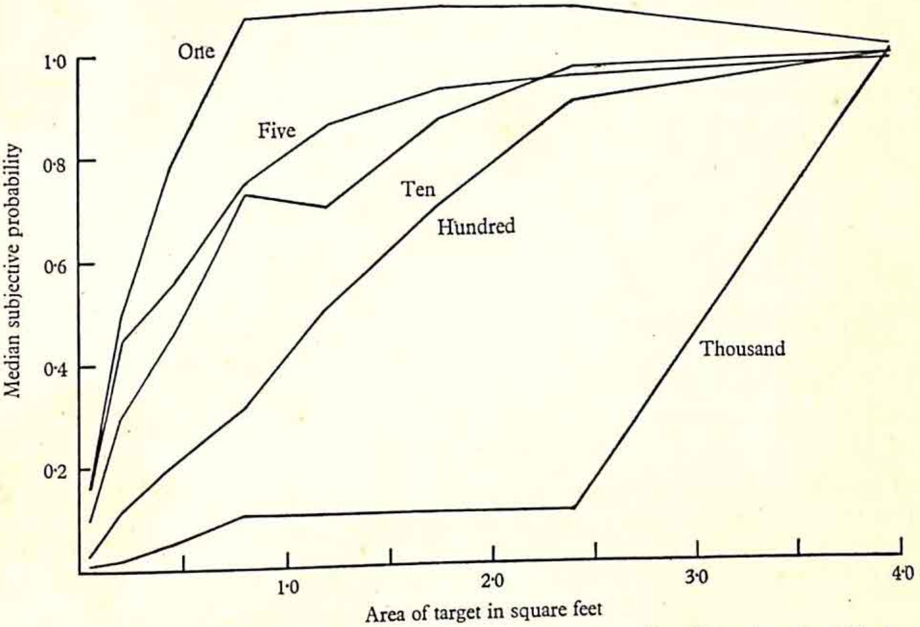


Fig. 1. Subjective probability of success in relation to area of target for different scales of judgement. ($n=62$ children age 9+).

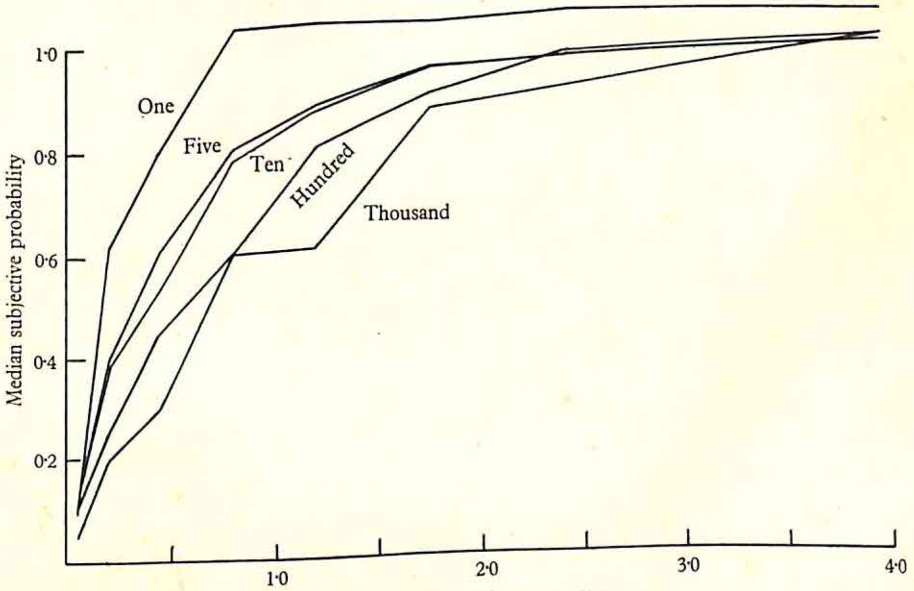


Fig. 2. Subjective probability of success in relation to area of target for different scales of judgement ($n=68$ children age 13+).

Table 2. *Median estimates on the 'hundred' scale corresponding to estimates on the 'five' scale for all targets*

	'Out of five'					
Age	0	1	2	3	4	5
9–10	2	8	20	30	35	90
13–14	8	15	30	45	75	99
'Expected' values	5	20	40	60	80	95

We are now in a position to give an affirmative answer to the question posed in the introduction above, as far as our subjects are concerned and under the conditions of the experiment. The fact that the range effect declines with age justifies the supposition that its effects would be slight in corresponding estimates made by adults. A further point of interest arose in an analysis of the distribution of estimates on the hundred scale. These show marked preferences for particular values varying with the age of the children. The values favoured by the younger group, after allowing for the differences in frequencies due to the difficulty of the task were 0, 1, 2, 3, 4, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 99 and 100. (A value is called 'favoured' if its corrected frequency exceeds 1% of the total frequency.) The older children favoured the values 1, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 70, 75, 80, 85, 90, 95, 99 and 100. These represented 70% of the values used by the younger children and 80% of the values by the older ones. The preferences suggest that in both age-groups the intervals of discrimination tend to be finer towards the ends of the range and coarser in the middle. This appears to indicate that ψ_2 values are high when they relate to either extreme of the ψ_1 scale and low when they refer to the middle region of the ψ_1 scale, as we found in experiments on intensity of belief (1956*a*).

The consistency of the individual estimates may be judged from a comparison of the medians of the two sets of values obtained from the target diameters (0·75 and 1·50 ft.) presented twice. These are shown in Table 3. The twenty pairs of medians (in Table 3) are fairly close to each other. We may safely assume that estimates of success at the other target diameters are equally reliable. We have also calculated the reliability coefficients between the sixteen pairs of estimates (omitting the 'out of one' range). Eleven of the sixteen coefficients exceed 0·7. In both age groups the majority of children chose 'out of one' as easiest, 'out of a thousand' as most difficult, and 'out of ten' as the range they would use to explain the difficulty of the task to a friend. The ψ_2 values, which signify

Table 3. *Median values of estimates of success at targets presented twice*

	Size of diameter							
Range	0·75 ft. Age				1·50 ft Age			
	9–10		13–14		9–10		13–14	
Out of:								
One	0·78	0·78	0·90	0·65	1·07	1·10	1·04	1·05
Five	3·3	2·4	3·8	2·9	4·6	4·7	4·8	4·8
Ten	4·5	4·7	5·9	5·2	8·5	8·8	9·4	9·6
Hundred	20	17	45	43	68	70	95	90
Thousand	101	31	328	300	106	155	894	800

an individual's degree of confidence in his estimates, range from 6·5 to 8·1. They differ significantly for the various ranges and show a slight tendency to decline as the range becomes larger. The older group give consistently larger values at each range.

IV. CONCLUSIONS

Our main conclusions may be briefly summarized as follows. (i) Subjective probabilities of success at a given task, as measured, tend to decline as the number of hypothetical attempts at the task become larger.* This effect is much less marked in the 13–14 age-group, than in the 9–10 age-group, and we may infer that it would be negligible in adult estimates. (ii) On all ranges, the children show preferences for certain values. These indicate a finer discrimination towards the ends of the range than in the middle region; the older children show a finer discrimination than the younger ones. A range 'out of a hundred' is warranted for the age groups 13–14, although they, like the younger children expressed a preference for the 'out of ten' range.

REFERENCES

COHEN, J. & HANSEL, C. E. M. (1956*a*). *Risk and Gambling*. London: Longmans Green.

COHEN, J. DEARNALEY, E. J. & HANSEL, C. E. M. (1956*b*). Risk and hazard. *Operat. Res. Quart.* **10**, 67–82.

(*Manuscript received* 5 *June* 1956)

---

* A similar effect is found in subjective probabilities of failure. The present paper should be read in conjunction with a subsequent study by the writers on the relationships between subjective probabilities of success and failure. (The Addition of Subjective Probabilities, *Acta Psychol.* 1956, **12**, 371–80.)

THE QUALITATIVE INVESTIGATION OF INTELLIGENCE

By ANN SHEDDEN McCOLSKEY

University of Florida

I. Introduction

A battery of seven intellectual tasks, designed for the qualitative examination of intelligence, was administered to thirty-five (experimental) children of superior ability and thirty-five (control) children of average ability, matched for age and other variables, and selected by a set of multiple criteria, including teachers' opinions, grades, and intelligence and achievement tests.

Under the conditions of this study, superior intelligence was characterized, with significant frequency, by a variety of qualities, including: coherence and productivity of informal thought processes; a facility for abstracting conceptual principles as against 'situation-bound' responses; conceptual flexibility; perseverance and independence of structure (or aid) in complex problems; attentiveness and spontaneous perception of incongruities. It is felt that this approach to intelligence can ultimately produce a more fertile, inductive quantification, in dimensional terms.

The undivided attention afforded the quantification of intellectual behaviour by psychologists since the early part of this century—reinforced by the lay enshrinement of the I.Q.—has unhappily stunted the growth of the logically complementary, analytic, or qualitative, approach. Practical exigencies fostered neglect of the controlled investigation of the more basic, complex, intellectual processes (which should *precede* quantification of an adequately complex and representative sort) and favoured the leap to immediate quantification of intelligence, on a more or less intuitive basis (cf. Wechsler, 1944, p. 13; Hearnshaw, 1951, p. 316), and in terms of the more superficial and palpable, intrinsically metrical aspects of intellectual functioning—speed, accuracy, retention, extent of vocabulary, and the like.

The inevitable result of intelligence testing of this sort is an array of paradoxes. The clinician's primary skill, for example, in reporting intelligence test performances, lies in his ability to resurrect, *intuitively*, the complex qualitative, or processual, context for the occasional intrinsically quantifiable response. Again, the expedient, *a priori* equation of intelligence with (palpable, intrinsically metrical) verbal skills (cf. Terman & Merrill, 1937, p. 5; Anastasi & Foley, 1954, *passim*) has given birth to a paradoxical environmental dependency-relationship which finds intelligence fluctuating according to cultural, subcultural, racial, regional, and rural-urban differences; deafness; foster home situations; and a host of other conditions, and fetters research by its inability to transcend such inequalities in opportunity for verbal development.

If one refuses the ultimate tautology of the traditional approach ('Intelligence is what the tests measure'), the alternative is a reorientation towards *representative* quantification, *inductively* derived from analytic observations of the complex mental processes which accompany the occasional, directly quantifiable response, and which fundamentally differentiate individual performances. The initial problem, which this study attacks, is one of devising a standardized framework for the observation and recording of basic behaviour patterns in a variety of intellectual tasks; of categorizing the results, and of determining significant differences in preference for the various categories—the qualitatively different modes—of intellectual operation between differing levels of intellectual ability.

II. Procedure

A qualitative battery. The battery devised for the present study (described in full elsewhere, McColskey, 1955) includes seven tasks, of varying length and complexity. Through each task is threaded a standardized administrative procedure for every conceivable level of response; this facilitates categorization, yet permits complete freedom of response. Each task also controls or by-passes the more usual, superficial measures of quantitative success, in order to minimize the effects of differential training and experience, as well as of motivational differences based on a consciousness of success or failure,* and to isolate the more 'native', covert, qualitative differences in behaviour between two levels of mental ability. No subject 'passes' or 'fails' a task in this battery.

(1) In a free-association 'Word-naming' task, a three-minute flow of words is recorded verbatim, under a set toward continuity, not speed (as in the Binet task), and is examined to determine whether significant differences in logical coherence complement the traditionally remarked differences in productivity, or the gross number of words.

(2) A tripartite 'Vigotsky' task employs the 22 blocks of the Vigotsky test, but departs considerably from the traditional administrative form. The basic Vigotsky problem requires the subject to sort the blocks, which comprise varying combinations of colour, shape, height, and surface-area, into four coherent groups. The difficult solution, which involves the interactive concepts of height and size, is realized through a trial-and-error mental process which is concretely reflected in the block manipulations. The traditional measures of time and of errors (Hanfmann & Kasanin, 1937) are lamentably sterile, as they are neither demonstrably reliable (Semeonoff & Laird, 1952, p. 99), nor representative of the stage-by-stage conceptual processes reflected by the test, as Heidbreder (1948) and others lament. The writer has constructed a systematic observational schema for the test, which, among other things, provides for presentation of successive series of four blocks of a shape, in a predetermined order. (Previously, clues have been offered in a hit-or-miss fashion, and an error of administrative judgment 'may alter the entire course of the subject's subsequent performance' (Semeonoff & Laird, 1952, p. 99).) The problem is thus progressively structured for each subject, until all achieve 'success'.

A similar sequence of progressive aid facilitates categorization of the verbal explanations of the correct groupings, and contrasts the initial, spontaneous, offerings with the levels reached with prompting. (Categorization of explanations was aided by suggestions from Hanfmann & Kasanin (1937); Ray (1936); Semeonoff & Laird (1952).) Two preliminary Vigotsky tasks also have been introduced, which first require the subject to sort the blocks on the simplest possible conceptual basis (by colour or by shape), and then to re-sort by an alternative concept (shape or colour). These transitional tasks provide a uniform frame of reference for the Vigotsky problem, and capture qualitative differences in *simple* conceptual operations, and in conceptual *flexibility*, to contrast with differences in the *complex* conceptual processes of the Vigotsky problem proper.

(3) Four short 'Absurdities' (adapted from Terman & Merrill, 1937) are presented in the guise of memory items, with discreet questioning to test for qualitative differences in the spontaneous appreciation of incongruities.

(4) Three simple 'arithmetic' problems, termed 'Confusions', are insoluble because of inadequate information, but are designed to mislead the unwary into offering a specious 'answer', and to search out qualitative differences in mental alertness or attentiveness.

(5) Three Rorschach cards are utilized to test for qualitative differences in modes of perceptual organization—in 'global', or unitary, 'whole' responses against differentiated 'whole' percepts—which would resemble certain age-changes reported in several Rorschach developmental studies (Hertz & Ebert, 1944; Meyer & Thompson, 1952; Stavrianos, 1942).

(6) Qualitative differences in 'Definitions' of nine, very simple, common words are examined, *apart from* the more usually investigated, culturally influenced, differences in *range* of vocabulary. (Categorization of 'Definitions' was facilitated by Feifel and Lorge's (1950) categorization of ontogenetic changes in quality of definitions.)

(7) Six pairs of simple, common words, selected from the Binet and Wechsler scales, comprise a 'Similarities and Differences' task, which aims to reveal qualitative differences in the complementary capacities for perceiving relationships and for differentiation.

The subjects. The battery was administered individually to 70 fourth- and fifth-grade children of two North Florida towns, in the summer of 1954. Thirty-five children were of superior intelligence; the other

---

* An after-inquiry of likes and dislikes revealed *unanimous* and enthusiastic endorsement of this 'new kind of test'.

35 constituted a control group, of average intellectual ability. In each group, there were 18 boys and 17 girls; 13 fifth-graders and 22 fourth-graders, and the socio-economic status of the average group approximated that of the superior group. Teachers' opinions were employed as preliminary criteria for selection; recent intelligence test scores, school marks, and achievement test scores over a 2 to 3-year period were then considered, *interdependently*. (In general, none of these crude ability indicators was given independent weight, but each child passed cutting points on all four criteria, or on three of four, if evidence on the fourth were borderline, and other data were extraordinarily impressive.)

The composition of each group is described fully elsewhere (McColskey, 1955), but may be sketched here briefly. On the California Test of Mental Maturity, which routinely had been administered in the schools, for example, superior I.Q.'s ranged from 125 to 147, with a mean of 134; comparable statistics for the average children were 90–110, and 102. Overlap was minimized but not prohibited, for scores along the remaining two objective criteria, in view of the relatively restricted range, and varying school and teacher standards; borderline cases here were assessed in terms of the consistency of all other data, as stated above. Thus, in honour-point equivalents, the superior group ranged from 3·0 to 4·0, around a mean of 3·7; corresponding data for the control group were 1·5 to 3·6, and 2·5. In achievement test scores, superior children ranged from 0·5 years below to 1·9 years above actual grade level, and were accelerated 1·0 year, on the average, while similar figures for the control group were 1·3 (below) to 0·6 (above), and 0·3 years' retardation.

III. RESULTS

Treatment of results consisted of categorization of varying behaviour-reactions and -patterns within each task, and comparison of experimental and control group frequency distributions by the χ^2 technique. (Marginal distributions provided the 'expected' proportions; neighbouring levels were combined as necessary, to ensure five or more 'expected' subjects; Yates' correction for continuity was employed where appropriate. It should be noted that the χ^2 treatment in the present paper represents a refinement of that employed in the original study.) Within any selected dimension of behaviour, categories are presumed to be exhaustive; are built on concrete referents, not subjective judgments; and are inductively derived from (not imposed on) the obtained protocols. Complete data are presented in the original study; a more succinct summary of findings is offered here.

Word-naming. A brief analysis of the logical coherence within protocols revealed significant differences ($P < 0{\cdot}02$) between experimental and control groups, in favour of the former. Coherence was defined in the simplest and most objective fashion possible, for this preliminary study, by classifying protocols according to the number of clear thought sequences (three or more obviously related words, comprising a generic sequence as: 'red, white, blue', or an associative train, as: 'lion, roar, zoo') as opposed to isolated or questionably related words. Group distributions may be illustrated by the fact that 40 % of the superior group and only 11 % of the average employed six or more sequences, while, at the other extreme, 31 % of the average protocols and 9 % of the superior were almost entirely disjunctive in quality, exhibiting two or fewer sequences. The expected group differences were found in the more commonly explored, complementary dimension of productivity (total number of words), with one-third of the superior group against 9 % of the average naming 60 or more words, and 26 % of the average against 0·0 % of the superior limited to 39 or fewer words ($P < 0{\cdot}01$). Further investigation (of mean coherence per productivity level, and vice versa) indicated that these twin aspects of thought processes in flux are less than perfectly correlated, especially among average children, so that a high productivity level does not necessarily imply notable coherence of thought.

Vigotsky, preliminary tasks. There were no group differences in the initial selection of a simple principle—colour or shape—for sorting the Vigotsky blocks, in (preliminary)

Task I: approximately two-thirds of both groups first chose shape as the preferred concept. Interesting and significant group differences in the initial reactions to the instructions for this relatively novel and primitive conceptual task were discovered, as indicated in Table 1. (Initial reactions were not systematically observed for all subjects on subsequent Vigotsky tasks. Statistics computed for those which were observed indicated increasing differentiation of groups with increasing task complexity, and a progressive downward shift in the locus of greatest differentiation.) Nearly one-third of the superior subjects impatiently began Task I before the examiner completed her instructions, while only one average subject did so. On the other hand, 43% of the average subjects demonstrated some difficulty in comprehension or acceptance of the task, in contrast to 14% of the experimental group.

Table 1. *Initial reactions to Vigotsky, preliminary tasks*

	Frequencies		
	Observed*		
Initial reaction	Av.	Sup.	Expected
A. Impatience: too-early start	1	11	6·0
B. Immediate start	19	19	19·0
(B). Immediate start, but questions during instructions	5	2	10·0
C. Delayed start (sits in thought, or asks several preliminary questions)	7	2	
D. No start; requires repetition of instructions	3	1	
Totals	35	35	35·0

* $\chi^2=13{\cdot}334$; $P<0{\cdot}01$ (D.F. $=2$: $P<0{\cdot}01$, $\chi^2=9{\cdot}210$; $P<0{\cdot}001$, $\chi^2=13{\cdot}82$).

Table 2. *Final (prompted) explanations of Vigotsky sortings, preliminary tasks*

	Frequencies		
	Observed*		
Final explanation	Av.	Sup.	Expected
A. Concept abstracted and named	25	33	29·0
B. Groups separately described, correctly	10	1	6·0
C. Irrelevant, vague, partial descriptions of groups	0	0	
D. Fails to explain	0	1	
Totals	35	35	35·0

* $\chi^2=4{\cdot}298$ (corrected for continuity); $P<0{\cdot}05$ (D.F. $=1$: $P<0{\cdot}05$, $\chi^2=3{\cdot}841$; $P<0{\cdot}02$, $\chi^2=5{\cdot}41$).

Classification of subjects according to the mode of grouping the blocks revealed insignificant differences in Task I, although 26% more superior than average subjects preferred an uncomplicated sorting into (5) colours or (6) shapes, while 26% more average than superior children produced a differentiated (one or more of the five or six primary groups subdivided in some way); an apparently confused; or no sorting.

Similarly, classification of the initial (unprompted) explanations of the (correct) sorting revealed no significant group differences. A slight initial superior proclivity (54% against 37%) for efficient abstraction and naming of the conceptual principle involved, however, did become significantly pronounced in the final explanations, with the admixture of systematic prompting by the examiner. As shown in Table 2, virtually all of the superior children were able to reach this level with prompting, while more than one-fourth of the control group remained 'situation-bound'—limited to description of the specific block-groups.

Upon instructions to shift to an alternative elementary sorting principle, in (preliminary) Task II, significant group differences in response again were demonstrated, as seen in Table 3. Nearly three-fourths of the superior youngsters, in contrast to approximately one-third of the average group, evinced independent conceptual flexibility, at this primitive level.

Table 3. *Shifting simple concepts, Vigotsky, preliminary tasks*

	Frequencies		
	Observed*		
Flexibility	Av.	Sup.	Expected
A. Spontaneous, appropriate re-sorting	12	25	21·0
B. Re-sorts with aid, after initial:			
B_1, selection of new, but inappropriate concept	2	3	
B_2, perseveration with original concept	6	2	6·5
B_3, vague, unclear sortings	5	0	
B_4, aimless or no activity	4	2	7·5
C. Rigidity: any of B_1–B_4, but no re-sorting, in spite of aid	6	3	
Totals	35	35	35·0

* $\chi^2 = 12{\cdot}562$, $P < 0{\cdot}01$ (2 D.F.: $P < 0{\cdot}01$, $\chi^2 = 9{\cdot}210$; $P < 0{\cdot}001$, $\chi^2 = 13{\cdot}82$).

Vigotsky, Task III. The Vigotsky problem proper provided some of the most intriguing and clear-cut group differences in the entire battery. Categorization of subjects by the level of structure required to achieve the correct Vigotsky sorting very significantly ($P < 0{\cdot}001$) differentiated the two groups. With no more than two levels of structure, for example, 74% of the superior children reached the solution, in contrast to 29% of the average subjects. Conversely, even with maximal structure (17 of the 22 blocks correctly placed for the subject), 37% of the average group against 3% of the superior still could not perceive the solution.

Group differences in the time taken to reach solution reaffirmed the marked superior tendency toward perseverance in an unstructured problem situation, and average inclinations toward inertia, with fairly constant demands for structure. Superior subjects took significantly *longer* than average ($P < 0{\cdot}05$) to reach solution, with 83% of the latter solving the problem in fifteen minutes or less, in contrast to 49% of the superior group. (These data cast some light upon the anomalous results of studies which have assumed speed of solution to constitute a positive measure of success in the test.)

Once all subjects had achieved (or been shown) successful solution of the Vigotsky problem, explanations of the correct sorting—both the initial, spontaneous offerings, and the final, prompted, verbalizations—still served to differentiate the two groups to a remarkable degree, as Table 4 demonstrates. Three-fourths of the superior children, in contrast to but 14% of the average, were able spontaneously to perceive and verbalize the twin concepts, in some form. Nearly one-half of the average subjects, against 11% of the superior, could not offer even a partially adequate explanation, without assistance. Yet the introduction of prompting exaggerated, rather than minimized, group differences in response, as 89% of the superior children achieved succinct abstraction of the twin concepts, or at least accurate group descriptions, with aid, while 77% of the average group still could not advance beyond successive, 'situation-bound', comparison of pairs of block-groups.

Table 4. *Initial and final explanations, Vigotsky*

Explanation	Initial (spontaneous) explanation (frequencies)			Final (prompted) explanation (frequencies)		
	Observed*			Observed†		
	Av.	Sup.	Expected	Av.	Sup.	Expected
A. Double concept abstracted and named ('It's by size and height')	1	3	15·5	2	24	13·0
B. Accurate description of each group, concepts used interactively ('These are big and tall; these are small and tall')	1	17		6	7	6·5
(B.) Successive comparison of groups, concepts used alternately ('These are taller than these, but bigger around than those')	3	6		5	0	7·0
C. Single concept abstracted and named ('It's by size')	11	5	10·0	5	4	
D. Description of each group, with variations on a single concept ('Big, medium, small, tiny')	4	0		9	0	8·5
E. Irrelevant, vague description of groups ('They're different shapes...')	7	3	9·5	3	0	
F. None offered	8	1		5	0	
Totals	35	35	35·0	35	35	35·0

* $\chi^2 = 25{\cdot}593$, $P < 0{\cdot}001$ (2 D.F.: $P < 0{\cdot}001$, $\chi^2 = 13{\cdot}82$).
† $\chi^2 = 38{\cdot}264$, $P < 0{\cdot}001$ (3 D.F.: $P < 0{\cdot}001$, $\chi^2 = 16{\cdot}27$).

Table 5. *Absurdities*

Recognition of incongruity	Item 1. Frequencies			Item 2. Frequencies			Item 3. Frequencies			Item 4. Frequencies		
	Observed*		Expected	Observed†		Expected	Observed‡		Expected	Observed§		Expected
	Av.	Sup.		Av.	Sup.		Av.	Sup.		Av.	Sup.	
A. Immediately precise	4	15	12·5	3	15	10·0	6	21	16·5	15	27	21·0
(A) Circumlocutory but, eventually correct	2	4		0	2		2	4		7	4	5·5
B. Possible vague notion, but never explicitly stated	4	3	22·5	0	0	25·0	0	0	18·5	6	0	8·5
C. None	25	13		32	18		27	10		7	4	
Totals	35	35	35·0	35	35	35·0	35	35	35·0	35	35	35·0

* $\chi^2 = 10{\cdot}516$, $P < 0{\cdot}01$ (1 D.F.: $P < 0{\cdot}01$, $\chi^2 = 6{\cdot}635$; $P < 0{\cdot}001$, $\chi^2 = 10{\cdot}83$).
† $\chi^2 = 13{\cdot}720$, $P < 0{\cdot}001$ (1 D.F.: $P < 0{\cdot}001$, $\chi^2 = 10{\cdot}83$).
‡ $\chi^2 = 16{\cdot}568$, $P < 0{\cdot}001$ (1 D.F.: $P < 0{\cdot}001$, $\chi^2 = 10{\cdot}83$).
§ $\chi^2 = 9{\cdot}010$, $P < 0{\cdot}02$ (2 D.F.: $P < 0{\cdot}02$, $\chi^2 = 7{\cdot}82$; $P < 0{\cdot}01$, $\chi^2 = 9{\cdot}210$).

After all subjects had offered or had been provided A-level explanations (at the conclusion of progressive, level-by-level, demonstration and discussion), instructions to demonstrate a firm grasp of the Vigotsky principle by re-sorting the blocks into the correct block-groups without aid still significantly differentiated experimental and control subjects ($P < 0{\cdot}05$), in terms of errors made. Sixty-nine percent of the superior children, for example, against 31% of the average, re-grouped without error.

Absurdities. Spontaneous appreciation of the incongruities involved in the four 'memory' items was consistently and significantly characteristic of the superior group, rather than the average, as Table 5 demonstrates. On all four items, superior children strongly predominated (by 32% or more) at a level of immediate sensitivity to the incongruity, while, on three of the four items, the large majority (71 % or more) of average children betrayed no recognition whatsoever. Summarizing these results, a re-categorization of subjects according to the number of 'Absurdities' explicitly recognized, immediately or eventually, very significantly differentiated the two groups ($P < 0{\cdot}001$), with 54 % of the experimental group and but 9 % of the control responsive to at least three of the four items.

Confusions. As can be seen in Table 6, reactions to two of the three *Confusions* significantly distinguished experimental and control groups, with the vast majority of the former (69 % or more) explicitly or implicitly alert to the subtle omissions in these 'arithmetic' problems, while the bulk of the latter (69 % or more) successfully were led astray. The first and most devious problem apparently deceived or confused both groups, although trends here were consistent with the pattern elsewhere. An overall, summary classification of subjects by the number of 'Confusions' resolved, explicitly or implicitly, significantly differentiated the two groups, with 60 % of superior children alert to at least two of the discrepancies, and 51 % of average children perceiving none.

Table 6. *Confusions*

	Item 1. Frequencies			Item 2. Frequencies			Item 3. Frequencies		
	Observed*		Expected	Observed†		Expected	Observed‡		Expected
Attentiveness	Av.	Sup.		Av.	Sup.		Av.	Sup.	
A. Refusal of solution, because of inadequate information	2	7	5·5	7	19	13·0	5	22	15·0
B. Solution, pending further information	2	0		3	5	6·5	1	2	
(B). Solution, supplying missing information	7	11	9·0	1	4		0	0	20·0
C. Deep thought; vague perception of discrepancy; no solution attempted	5	7	6·0	0	0	15·5	2	0	
D. Immediate false solution	15	10	14·5	24	7		27	11	
E. Blank; total confusion; immediate refusal of problem ('I'm no good at arithmetic')	4	0		0	0		0	0	
Totals	35	35	35·0	35	35	35·0	35	35	35·0

* $\chi^2 = 4{\cdot}834$, N.S.
† $\chi^2 = 16{\cdot}784$, $P < 0{\cdot}001$ (2 D.F.: $P < 0{\cdot}001$, $\chi^2 = 13{\cdot}82$).
‡ $\chi^2 = 18{\cdot}900$, $P < 0{\cdot}001$ (1 D.F.: $P < 0{\cdot}001$, $\chi^2 = 10{\cdot}83$).

Perceptual organization. Qualitative differences in the mode of organization of 'whole' Rorschach responses were significant on Card 1 ($P < 0{\cdot}01$), but not on Cards 2 and 7, although the direction of differences was the same in all three cases. (The latter two cards produced very few 'global' responses, in either group.) On Card 1, approximately three-fourths of the average subjects, against 36 % of the superior, reported a 'global' 'whole' percept only, while 64 % of the superior children, against 26 % of the average, responded either with a differentiated 'whole' or with *both* differentiated *and* 'global' 'whole' percepts.

Definitions. Of the five 'Definitions' ('bicycle', 'puddle', 'nonsense', 'join', and 'compare'), in which the factor of range of vocabulary was excluded as desired (i.e. in which all subjects were thoroughly familiar with the words), in order to isolate more or less 'native', qualitative definitional differences, four significantly differentiated experimental and control groups ($P < 0{\cdot}001$ in three cases; $< 0{\cdot}01$ in one), and all five items demonstrated the same pattern of differences. Categorization was, of course, specific to the 'Definition', but a typical overall progression emerged, which is illustrated in Table 7. In all five cases, superior subjects most often (from 44 % to 91 %) offered either concise synonyms, or fuller descriptive explanations, while average definitions most typically (from 45 % to 70 %) were either overly vague and general or overly narrow and particular, or else wholly amorphous.

Similarities and differences. No formal results in this task can be presented because the fact that several subjects offered more than one response (χ^2 demands independence of frequencies in each class) precluded computation of probability figures, for group differences. In general, group differences in response were marked—somewhat more so in 'Similarities' than in 'Differences'—and resembled those found in 'Definitions'. The extraordinary heterogeneity of the item-categorizations in this task, however, defied extraction of a clear-cut overall progression, although superior responses tended to comprise precise generalizations of basic similarities or differences, while average responses tended towards indefiniteness, irrelevance, or narrow particularization. In response to the first 'Similarity', 'car-boat', for example, 83 % of the superior and 42 % of the average responses consisted of the precise generalization, 'transportation'. The remaining 17% superior and 58 % average responses were more specific ('motors in both'); or vaguer ('both go'); or clearly irrelevant ('people made both').

Table 7. *Definitions of 'puddle'*

	Frequencies		
	Observed*		
Definition	Av.	Sup.	Expected
A. Precise synonym ('standing water')	6	14	10·0
B. Descriptive explanation ('a place where rain makes mud in the road')	9	18	13·5
C. Vague characterization; situational illustration ('some water'; 'puddles are in back of my house after rain')	6	1	11·5
D. Repetitive or inadequate phrases ('a puddle of water'; 'it rains')	14	2	
Totals	35	35	35·0

* $\chi^2 = 18{\cdot}766$, $P < 0{\cdot}001$ (2 D.F.: $P < 0{\cdot}001$, $\chi^2 = 13{\cdot}82$).

IV. DISCUSSION AND CONCLUSIONS

It should be stressed that the limitations in scope of this brief study demand tentative rather than definitive conclusions. Among other things, there are (as yet) no reliability figures. Inter-observer reliability was very high (0·997) in an early study which utilized the qualitative approach (Terman & Merrill, 1937, p. 303, n.), and hence might be so, by analogy, in the present study, especially since this investigator was restricted to non-judgmental recording of behaviour. The Vigotsky presents the greatest barrier to an estimate of task reliability, in terms of parallel forms (some tasks bear simple retesting; the split-half technique appears less appropriate to a heterogeneous battery), but one study reports some progress in this direction (Semeonoff & Laird, 1952), and a gratifying stability of individual behaviour patterns.

It also should be kept in mind that the investigator both selected and tested the subjects. In order to prohibit any subliminal 'halo' effect, all identifying data for the subjects were filed away 3 months prior to testing, and protocols were coded and well-shuffled, after testing, with another 3 months' hiatus before categorization. Furthermore, the rigid adherence to recording and sorting in terms of concrete referents virtually eliminated any opportunity for subjective assessment.

On the other hand, this latter restriction also limited the scope of the investigation. The extension of observations and analysis through the use of observer-teams, as well as recording devices, would permit study of endless nuances of behaviour glossed over in this initial survey. In the 'Word-naming' task, for instance, one could explore differential trends in a more subtle, *overall* coherence of thought processes, and in the variability, extensiveness, and generality of content categories. In the Vigotsky tasks, initial reactions call for further study, along with the flexibility and relevance of presolution hypotheses, and working attitudes (labile or intensive concentration; preference for or avoidance of overt manipulation; self-sufficiency, etc.). In the Rorschach situation, the study of differential organizations of a *constant* area may be extended to major details, as well as wholes, with the elaboration and precision of a response (*within* the $F+$ to $F-$ dichotomy) calling for consideration, among other qualities.

Meanwhile, certain tentative conclusions, relevant to this particular procedure and normative group, may be offered. (1) The mental processes of superior children were significantly more coherent, as well as more vigorous or productive, than those of average children. (2) Subtle behaviour reactions to an elementary task significantly differentiated intellectual levels, while grosser task achievements did not. (3) Superior children significantly more often abstracted and named conceptual principles, while average children remained 'situation-bound'. (4) Superior children demonstrated a significantly lesser degree of conceptual rigidity than average children. (5) Superior children required significantly less structure than average (and persevered significantly longer without aid) in a complex conceptual problem, and they made fuller use of assistance, when given. (6) Significantly more superior than average children perceived and manipulated multiple concepts. (7) Superior children had a significantly firmer grasp than average on demonstrated conceptual principles. (8) Logical incongruities were spontaneously recognized significantly more often by superior than by average youngsters. (9) Superior children were significantly more alert or attentive to camouflaged omissions and discrepancies. (10) Average children showed a significant restrictive trend towards 'global' percepts, while superior youngsters added or preferred differentiated percepts. (11) Superior children tended to define even the most familiar and mundane words with synonyms or comprehensive descriptions, in significant contrast to average indefiniteness or over-specificity.

Apart from the specific findings, it is hoped that this paper accomplishes its broadest purpose, of stimulating a novel orientation toward the analysis and subsequent measurement of the more fundamental intellectual processes. The battery itself, of course, may be revised and extended in an endless variety of ways, and applied to many other subject-groups, both normal and abnormal, toward the gradual accretion of a body of knowledge —a set of qualitative norms—of differential intellectual processes, at various life-stages, and toward truly fertile, experimentally based, dimensional quantification of intelligence.

This paper is based on a thesis offered in partial fulfilment of the requirements for the M.A. degree at the University of Florida, June 1955. The author wishes to express her appreciation to Dr J. C. Dixon and Dr R. J. Anderson for their unflagging interest and stimulating criticisms.

REFERENCES

AMES, L. B., LEARNED, J., METRAUX, R. W. & WALKER, R. N. (1952). *Child Rorschach Responses: Developmental Trends from Two to Ten Years.* New York: Hoeber.

ANASTASI, A. & FOLEY, J. P. (1954). *Differential Psychology.* (Rev. ed.) New York: Macmillan.

BINET, A. & SIMON, T. (1916). *The Development of Intelligence in Children.* Baltimore: Williams and Wilkins.

FEIFEL, H. & LORGE, I. (1950). Qualitative differences in the vocabulary responses of children. *J. educ. Psychol.* **41**, 1–81.

HANFMANN, E. & KASANIN, J. (1937). A method for the study of concept-formation. *J. Psychol.* **3**, 521–40.

HEARNSHAW, L. S. (1951). Exploring the intellect. *Brit. J. Psychol.* **42**, 315–21.

HEIDBREDER, E. F. (1948). Studying human thinking. In T. G. Andrews (ed.), *Methods of Psychology.* New York: Wiley.

HERTZ, M. R. (1951). *Frequency Tables for Scoring Responses to the Rorschach Ink Blot Test.* Cleveland: Western Reserve University Press.

HERTZ, M. R. & EBERT, E. H. (1944). The mental procedure of 6 and 8 year old children as revealed by the Rorschach ink-blot method. *Rorschach Res. Exch.* **8**, 10–30.

McCOLSKEY, A. S. (1955). *An Investigation of Qualitative Differences in Intelligence between Average and Superior Children.* Unpublished Master's Thesis, University of Florida.

MEYER, G. & THOMPSON, J. (1952). The performance of kindergarten children on the Rorschach test: a normative study. *J. Proj. Tech.* **16**, 86–111.

RAY, J. J. (1936). The generalizing ability of dull, bright and superior children. *Peabody Contr. Educ.* No. 175.

RICHMOND, W. K. (1953). Educational measurement: its scope and limitations. A critique. *Brit. J. Psychol.* **44**, 221–31.

SEMEONOFF, B. & LAIRD, H. J. (1952). The Vigotsky test as a measure of intelligence. *Brit. J. Psychol.* **43**, 94–102.

STAVRIANOS, B. (1942). An investigation of sex differences in children as revealed by the Rorschach method. *Rorschach Res. Exch.* **6**, 168–75.

TERMAN, L. M. & MERRILL, M. A. (1937). *Measuring Intelligence.* Boston: Houghton Mifflin.

WECHSLER, D. (1944). *The Measurement of Adult Intelligence.* (3rd ed.) Baltimore: Williams and Wilkins.

WECHSLER, D. (1949). *Wechsler Intelligence Scale for Children.* New York: Psychol. Corp.

(*Manuscript received* 25 *September* 1956)

THE SOLVING OF THREE-TERM SERIES PROBLEMS

By IAN M. L. HUNTER

Psychology Department, Edinburgh University

Sixteen problems of the three-term series type are presented to each of 64 11-year-old children and 32 16-year-olds. The older children show evidence of solving these problems by articulated sequences of analytical judgments which accord with the varying logical structures of the problems, show no influence of atmosphere effect, show marked practice effects in progressing from one problem to the next, and solve problems involving the relatively abstract 'happier-sadder' relation as quickly as identically structured problems involving the relatively concrete 'taller-shorter' relation. In contrast, the younger children show but slight evidence of being influenced differentially by variations in logical structure, are susceptible to atmosphere effect, show smaller practice effects, and solve 'taller-shorter' problems more rapidly than 'happier-sadder' problems. The results are interpreted as demonstrating that increasing age brings increasing appreciation of the structural characteristics of series as such, together with increasing skill in dealing with serial relations which are progressively more remote from the perceptual-motor level of behaviour.

I. Introduction

'Tom is taller than Dick: John is shorter than Dick. Which of these three boys is the tallest?' This exemplifies what is here called a three-term series problem. Consider the series $A > B > C$. This series may be transformed into a set of two premises, each premise comprising two terms and a serial relation. (For a discussion of the logical properties of series see, for example, Stebbing, 1933, pp. 172ff.) When such a set of premises is presented to someone who is required to derive a three-term series from it, then the person is said to have a three-term series problem. This class of problem is not completely representative of series problems at large since it is restricted on three counts. First, the items to be seriated are not presented in concrete actuality as, for example, in Piaget's experiments with sticks, dolls, etc. (1952, pp. 96–157); rather they are presented verbally. Secondly, the series to be derived comprises the minimum possible number of items, namely, three; this contrasts with the four- and five-term series employed, for example, by Pinard, Barbeau, Laurendeau & Parant (1954) in their adult intelligence test. And thirdly, the problems, as stated, make no use either of the 'between' relation or of negative statements such as 'A is not taller than B'. Nevertheless, the very restrictedness of this class of problems makes for easier experimental investigation of at least some of those psychological processes which would seem to be fundamental to the solving of any type of series problem. The central question is: by what psychological processes do people derive the series from the given set of premises? Some part of an answer has been offered by Piaget (1921, 1928) but by far the greatest amount of data bearing on this question is furnished by Burt (1919). The discussions of Burt, with his predominantly statistical approach, and of Piaget, with his predominantly 'clinical' approach, suggest a number of variables likely to affect the solving of three-term series problems. It is the purpose of the experiment reported below to examine some of these variables further. But before describing this experiment, it may be best to clarify some points of terminology and present a theoretical outline of the solving process.

(i) *The structure of three-term series problems*

Consider the statement that A is taller than B who is taller than C. This series may be represented symbolically as $A>B>C$ or as $C<B<A$. It may also be presented as a set of two premises, the four basically different premise structures being as follows: I, $A>B:B>C$; II, $A>B:C<B$; III, $B<A:C<B$; IV, $B<A:B>C$. A, B, and C are the *terms* of the premises and likewise of the series. Any given serial relation is represented by $>$ (or by $<$) while $<$ (or $>$) represents the same relation *reversed*. The *converse* of any premise is that proposition in which the terms are transposed and the relation reversed, e.g. $B>C$ is the converse of $C<B$. The *referent* of any premise is its first term and the *referee* is its second term, e.g. A is the referent of $A>B$ and the referee of $B<A$. Two premises are *linked* where they have one term in common, e.g. $A>B:B>C$ are linked but $A>B:C>D$ are not. Two linked premises are *isotropic* when each contains the same unreversed relation and also when the common term is the referee of the first-encountered premise and the referent of the second premise, e.g. $A>B:B>C$ and $C<B:B<A$. When the relation of one linked premise is the reverse of that contained in the other, then these premises are *heterotropic* e.g. $A>B:C<B$. Likewise, when the relations are identical but the common term is the referent of the first premise and the referee of the second, then the linked premises are, again, heterotropic, e.g. $B<A:C<B$. (Notice that, since only two premises are involved, they will necessarily be encountered consecutively. With problems involving series of four or more terms, linked premises may be non-consecutive, e.g. $A>B:C>D:B>C$. In such a case, the person must, after encountering the first premise, suspend judgment, and either wait or search for that second premise which is linked to it. Examples of problems made more difficult by the non-consecutive presentation of linked premises are to be found in the relational sorites of Lewis Carroll (1895), and in the series problems of Pinard *et al.* (1954).)

Table 1

	First set		Second set	
I, 1	$A>B:B>C$	$>?$	$C<B:B<A$	$<?$
I, 2	$A>B:B>C$	$<?$	$C<B:B<A$	$>?$
II, 1	$A>B:C<B$	$>?$	$C<B:A>B$	$<?$
II, 2	$A>B:C<B$	$<?$	$C<B:A>B$	$>?$
III, 1	$B<A:C<B$	$>?$	$B>C:A>B$	$<?$
III, 2	$B<A:C<B$	$<?$	$B>C:A>B$	$>?$
IV, 1	$B<A:B>C$	$>?$	$B>C:B<A$	$<?$
IV, 2	$B<A:B>C$	$<?$	$B>C:B<A$	$>?$

The premises having been presented to the person, he may be required to derive the series by being asked one of several different questions. In the present study, only one form of question is employed, namely, 'Who is tallest/shortest?' In other words, the question requires the person to state that term which stands to the other two in the asymmetrical relation expressed by the superlative adjective (e.g. is tallest of) instead of by the comparative (e.g. is taller than). Thus each of the above four premise structures may be followed by one of two questions, that is, either by (1) 'Who is tallest?' which may be symbolized as '$>?$', or by (2) 'Who is shortest?' which may be symbolized as '$<?$' The eight different problem structures which result are shown in the first column of Table 1 preceded by a 'code reference' in terms of which each problem structure will be

referred to hereafter. The second column shows a further eight structures which are identical with the first eight except that, in each structure, there is a reversal of all the relations involved and a corresponding change in the locations of the terms. Since the ensuing discussion applies equally well to both alternative sets of eight structures, reference will be made only to the structures listed in the first column. (A complete study of three-term series problems involves considering a total of sixteen different structures; and these may be derived by arguments other than that outlined above. However, as Table 1 shows, half of these structures are, as it were, psychologically equivalent to the other half. Starting with the series $C > B > A$ instead of $A > B > C$, a further set of sixteen structures may be derived which are, in effect, identical with those shown in Table 1.)

(ii) *Reorganizings involved in deriving the series*

The covert psychological performances presumed to be involved in deriving a series from a set of two-term premises are here inferred on the basis of two assumptions. The first is that seriating necessitates the successive judging of premises which are not only linked (this is a logically necessary condition) but also isotropic. That is, they must be encountered in that sequence which James (1891), in discussing his 'axiom of skipped intermediaries', described by saying that 'any number of intermediaries may be expunged without obliging us to alter anything in what remains written' (p. 646). If two linked premises are not isotropic, the thinker must therefore make them isotropic by reorganizing one or other of the premises. The second assumption is that, where reorganizing is required, it is the second-encountered premise which tends to be reorganized rather than the first. That is, the first premise establishes the direction of the series, and predisposes the thinker to make the second premise isotropic with the first. Now, as regards the reorganizing performance required to make the two linked premises isotropic, this may (in optimally efficient reasoning) take one of three distinguishable forms according as to whether the problem is of the structure designated above as II, III, or IV. With II ($A > B : C < B$), the second premise requires simple *converting* to give $A > B : B > C$. With III ($B < A : C < B$), the second premise must be 'brought forward' or *reordered* to give $C < B : B < A$. With IV ($B < A : B > C$), isotropy may be achieved by two alternative procedures. In one of these, the second premise is both converted and reordered. Thus, $B < A : B > C$ becomes $B < A : C < B$ and then $C < B : B < A$. In the alternative procedure, it is the first-encountered premise which is converted. Thus, $B < A : B > C$ becomes $A > B : B > C$. But in order to achieve this converting, it is necessary to back-track on the first premise and reverse the direction established by this premise. Both of these alternative procedures involve, therefore, converting as well as some additional reorganizing. In the *converting-reordering* performance, it is the reordering which is additional. And in the *reversing-converting* performance, there is an additional reversing, that is, a 'going back' to and a reconsidering of the already encountered first premise and a reversing of the direction which has already been established by it. It is of interest that this reversing process has been singled out by Piaget in his discussions of 'operations', and by animal psychologists in discussing the 'reverse detour' as 'a back-and-forth movement inconsistent with one's natural expectation of a straightforward progress toward the goal' (Woodworth & Schlosberg, 1954, p. 834).

(iii) *Atmosphere effects*

The above discussion treats the deriving of a three-term series as a sequence of analytical judgments of the relation and identity between successively encountered terms. But both Burt and Piaget found that, with children, the deriving may be strongly influenced by the 'general feel' or 'global impression' of the problem: that certain forms of problem predispose the thinker to accept that particular answer which 'looks right' although it may or may not be correct. This influence of 'global impression' has received its most unambiguous confirmation in the study of adult syllogistic reasoning and has been called, by Woodworth & Sells (1935), the 'atmosphere effect'. Another and particularly compelling instance of an 'atmosphere effect' which derives from the structural characteristics of the problem has been reported recently by Hunter (1957).

From Burt's discussion of three-term series problems, it would appear that atmosphere may derive from two closely related sources. These may be called 'direct statement' and 'inclusiveness', respectively. The effect of 'direct statement' is described by Burt as follows. 'With the statement "Jack runs slower..." ringing in the memory, a child asked "who is the slowest?" naturally tends to say "Jack runs slowest"' (p. 126). This source of suggestion will presumably operate most strongly in structure II where one premise contains the given relation and the other the reversed relation and where there is no conflict between the two direct statements. Consider the problem 'A is taller than B: C is shorter than B. Who is tallest?' Here, the answer suggested by direct statement is the referent of the first premise, i.e. A. In a sense, the thinker is actually told that A is the tallest. Likewise, the referent of the second premise is suggested in answer to the question 'Who is shortest?' Therefore, irrespective of the question asked, problems of structure II suggest the correct answer by 'direct statement'. In contrast, problems of structure IV would suggest, by 'direct statement', the wrong answer, namely, the middle term. In this case, however, the two direct statements conflict, B being taller in one premise and shorter in the other. In consequence, it is doubtful whether atmosphere effect deriving from 'direct statement' would operate in structure IV.

'Inclusiveness' has been discussed by Piaget (1921) in relation to the following problem of structure IV, 2. 'Edith is fairer than Suzanne: Edith is darker than Lili. Which is the darkest, Edith, Suzanne, or Lili?' As Burt points out, the statement 'Edith is darker than Lili' suggests that both Edith and Lili belong to the dark group, Edith being merely more intensely marked by the special character of that group; hence, the child is apt to select Lili as the dark one when asked if she is darker than Suzanne. In Piaget's words, it is as though the child reasoned as follows. 'Edith is fairer than Suzanne, so they are both fair; Edith is darker than Lili, so they are both dark: therefore Lili is dark, Suzanne is fair, and Edith is between the two. In other words, owing to the interplay of the relations involved in the test, the child, by substituting the judgment of membership (Edith and Suzanne are "fair" etc.) for the judgment of relation (Edith is "fairer than" Suzanne), comes to a conclusion which is exactly the opposite of [the logical conclusion]' (1928, pp. 87–8). (It should be mentioned that the conclusions drawn by Piaget from his 1921 study are somewhat vitiated not only by his 'clinical method' but also by his use, in rephrasing the problems, of the words *blonde* and *brune* alongside the more relational words *claire* and *foncée*.) In the problem structures outlined above, 'inclusiveness' would favour an incorrect answer with IV and a correct answer with II.

In summary, if the two atmosphere effects were to operate, structure II would be easier because of both 'direct statement' and 'inclusiveness' while structure IV would be more difficult to solve because of 'inclusiveness' but not, presumably, of 'direct statement'.

(iv) *Reorganizing and ease of solving*

The processes suggested above as operating in the solving of three-term series problems are, at best, directly observable only to a limited extent through retrospecting and, at worst, not amenable at all to direct observation. However, if the ease with which any particular problem is solved were a function of the number and complexity of the reorganizings required, then these reorganizings would manifest themselves in the relative difficulties of each problem structure. Problems containing heterotropic linked premises would be more difficult than problems of structure I where the premises are already isotropic: IV would be more difficult than problems requiring only converting (II) or reordering (III): and if atmosphere effects operated, the negative atmopshere problems (IV) would be more difficult than the positive (II). Such differences in relative difficulty would reveal themselves either in the frequency with which the problems were solved under standard conditions or in the time required for solving, it being supposed that every additional reorganizing process increases both the likelihood of error and the time required to arrive at the correct solution.

II. PROCEDURE

(i) *Subjects*

There were two groups of subjects. The first comprised 64 children (32 boys and 32 girls) aged between 10 years 5 months and 11 years 11 months, with a mean age of 11 years 2 months. These children were drawn from two classes of the Moray House Demonstration School, Edinburgh, and were, as a group, above average intelligence. The second group comprised 32 children (14 boys and 18 girls) aged between 15 years 1 month and 16 years 11 months, with a mean age of 16 years 1 month. These children were also above average intelligence, being prospective candidates for the Scottish School Leaving Certificate. They all attended Portree Senior Secondary School, Isle of Skye. The reason for having twice the number of 11-year-old subjects will become apparent when the results are considered. Briefly, such small inter-structure differences emerged from the first 32 11-year-olds studied that it was thought advisable to repeat the whole experiment with a further group of 32 subjects.

(ii) *The problems used*

The sixteen problems used are shown schematically in Table 2, numbered 1 to 16. In the third column of that table, the structure of each problem is indicated by the 'code reference' described in the Introduction. The exact wordings of problem nos. 1, 2, and 3 were as follows. 'George is taller than Harry: Harry is taller than Willie. Which of these three boys is the tallest?' 'Lily is sadder than Olive: Lily is happier than Edith. Which of these three girls is the saddest?' 'Tom is taller than Dick: John is shorter than Dick. Which of these three boys is the tallest?' Each succeeding problem was worded in exactly the same way, the terms in each being either three short and familiar girls' names or three short and familiar boys' names. The specific names used were varied from problem to problem to avoid confusion; and the problems containing girls' names alternated with those containing boys' names. Throughout the sixteen problems, each of the eight different structures listed in the first column of Table 1 occurred twice, once with the 'taller-shorter' relation (these problems are designated by asterisks in Table 2) and once with an alternative serial relation. For the 16-year-olds and half of the 11-year-olds, this alternative was the 'happier-sadder' relation. 'Happier than' was substituted for the symbol '>' and 'sadder than' for the symbol '<'. For the remaining half of the 11-year-olds, 'warmer-colder' was the alternative relation: 'warmer than' was substituted for '>' and 'colder than' for '<'. Thus each of the 32 16-year-olds tackled eight problems involving the 'taller-shorter' relation interspersed among eight problems, of identical structure, involving the 'happier-sadder' relation. Each of 32 11-year-olds tackled exactly the same problems as the members of the older group. And each of the remaining 32 11-year-olds tackled sixteen problems of which eight involved the 'taller-shorter' and eight the 'warmer-colder' relation. Each problem was typed on a separate white 6 in. by 4 in. card. The first premise occupied the first line, the second premise a second line directly beneath the first, and the question a third line beneath the other two.

(iii) *Administration*

The procedure for both groups of children was identical. The children were tested individually in a room of the school attended by them. Each was given twenty problems in all, that is, four practice problems plus the sixteen test problems. When the child was seated and rapport established, he was instructed as follows. 'Now, we're going to do some simple problems and I think you'll find them easy enough. Here's the first one. Just read it over to yourself.' The first practice problem was handed to the child. 'Now, you see the question there at the end? Tell me what you think the answer is.' When this first problem had been correctly answered, the card was removed and the experimenter continued. 'Here's another problem much the same as before. But this time I'm going to time you with this watch to see how long you take. Tell me the answer as quickly as you can without, of course, hurrying so that you make mistakes. And if you do make a mistake, I'll tell you so that you can have another try.' With this, the experimenter presented the second practice problem, and started a stop-watch when the child began to read the problem. If a wrong answer was given, the experimenter said 'No, try again', and recorded the wrong answer along with the time at which it was given. The final (and usually the only) time recorded was that at which the correct answer was given. These times were entered in a prepared record blank to the nearest second; and the timing was carried out as unobtrusively as possible, without attempt to conceal from the child the fact of his being timed. Testing continued in this manner with the remaining practice problems and the sixteen test problems. The four practice problems were taken from Burt's (1919) graded series. They were tests nos. 2, 3, 8, and 11 of that series and were always presented in that order. The test problems were presented to the first sixteen children in the same cyclical order (i.e. nos. 1, 2, ..., 16) but each child began with a different problem. Likewise, each of the second sixteen subjects began with a different problem, the problems now being in the constant cyclical order of nos. 16, 15, ..., 1. Taking the group as a whole, each problem was thus tackled first, second, third, and so on by an equal number of children with the effect of counterbalancing the influences of possible practice effects on any particular problem. At no point was the child questioned about his 'reasons' for giving any particular answer. Occasionally, however, the child would spontaneously offer some explanatory remark; and, when this happened, the remark was recorded.

(iv) *Treatment of results*

Because, on the average, the 16-year-olds solved the problems more quickly than the 11-year-olds, the results for each group had to be treated separately. The general treatment of results was, however, identical for both groups. The basic assumption underlying the design of this study was that the introduction of an additional 'step' in the thinker's thought processes would involve an increase in the time taken to arrive at the correct solution. And the question was: did this problem take relatively longer to solve than that? For each problem there was, in each group, a distribution of either 64 or 32 time scores and the interest lay in comparing these distributions as between one problem and another. This comparison could not be made by t-tests because the distributions were not only skewed but also discontinuous as a consequence of some children having given a wrong answer before reconsidering the problem at the experimenter's prompting and arriving at the correct solution. The use of χ^2 was, however, appropriate. Accordingly, the median time for solving was found. This was between 11 and 12 sec. for the younger group and between 8 and 9 sec. for the older group. The total was then found of the number of instances in which a particular problem was solved correctly in less than the median time, and, likewise, of the number of instances in which that problem was not solved correctly in under the median time. However, it was discovered that this procedure threw away some of the data. Thus, one subject might have solved all sixteen problems in less than (or more than) the median time for the group; so, whatever time differences he showed between the solving of one problem and the next would not enter into the final results. The following modified procedure was, therefore, adopted. Instead of selecting the median time score for the total group of 32 or 64 children, each child was made to provide his own median value. This median varied from child to child but, for each individual, there were eight problems falling below and eight above (allowance being made for ties between problems). The time scores of each child were thus treated separately, and a note made of those problems which were correctly solved in the eight shortest as contrasted with the eight longest times. The total was then found, for all the children in the group, of the number of instances in which each problem was (*a*) one of the eight easiest to solve, and (*b*) one of the eight most difficult. (One point about this method merits mention. It is the treatment of those instances in which a problem was correctly solved in less than the median time even after a wrong answer had been given and corrected. [These solutions were regarded as falling below the median time.] This treatment is admittedly questionable; but in fact it was of no practical consequence since such ambiguous instances occurred in only 5 out of the total of 1536 instances.) The results were worked out by both the above methods, that is, by taking the group median and the individual medians. Exactly the same inter-problem differences emerged from each method; but the latter procedure proved its merit by revealing these differences in slightly enhanced form. It is the totals obtained by the latter (individual medians) method which are shown in Table 2 to two decimal places.

Table 2

No.	Structure	Ref.	11-year-olds, $N=64$. No. of solutions		16-year-olds, $N=32$. No. of solutions	
			Below median	Above median	Below median	Above median
1*	$A>B:B>C \quad >?$	I, 1	37·02	26·98	20·49	11·51
2	$B<A:B>C \quad <?$	IV, 2	22·12	41·88	6·32	25·68
3*	$A>B:C<B \quad >?$	II, 1	44·00	20·00	19·10	12·90
4	$B<A:C<B \quad <?$	III, 2	34·35	29·65	20·57	11·43
5*	$B<A:B>C \quad >?$	IV, 1	33·20	30·80	14·30	17·70
6	$A>B:B>C \quad <?$	I, 2	28·83	35·17	24·80	7·20
7*	$B<A:C<B \quad >?$	III, 1	34·58	29·42	15·25	16·75
8	$A>B:C<B \quad <?$	II, 2	34·47	29·53	17·30	14·70
9	$A>B:B>C \quad >?$	I, 1	33·02	30·98	25·62	6·38
10*	$B<A:B>C \quad <?$	IV, 2	25·35	38·65	5·05	26·95
11	$A>B:C<B \quad >?$	II, 1	33·10	30·90	15·01	16·99
12*	$B<A:C<B \quad <?$	III, 2	30·10	33·90	14·12	17·88
13	$B<A:B>C \quad >?$	IV, 1	24·35	39·65	7·93	24·07
14*	$A>B:B>C \quad <?$	I, 2	34·43	29·57	20·06	11·94
15	$B<A:C<B \quad >?$	III, 1	19·00	45·00	5·14	26·86
16*	$A>B:C<B \quad <?$	II, 2	45·33	18·67	23·43	8·57

* Denotes those problems which involve the 'taller-shorter' relation.

III. Results

(i) *Effect of practice*

People who perform a succession of similar tasks usually show progressive improvement from one task to the next. Do the children in this experiment show such an improvement? This question is answered by comparing the times taken to solve those problems which are tackled early in the sequence of sixteen with the times for those tackled late. With the 11-year-olds, the first four problems are solved in less than the median time in 115 (out of 256) instances, while the last four are similarly solved in 147 instances. The difference between these two proportions is significant at the 0·01 level ($\chi^2=8{\cdot}005$). With the 16-year-olds, the first four problems are solved in less than the median time in 58 (out of 128) instances and the four problems tackled last in 88 instances. This difference is significant at well beyond the 0·01 level ($\chi^2=14{\cdot}346$). Especially in view of the fact that there is only half the number of subjects in the older group, this difference is significant at a level which so far exceeds 0·01 as to lead to the conclusion that both groups show a practice effect, the older children much more than the younger.

(The same conclusion emerges when the first and last two or three problems—or even the first and last problems only—are compared.)

(ii) *Effect of premise structure*

The sixteen test problems contain only four different premise structures. These are shown in Table 3, together with the appropriate totals (to the nearest whole number) for both groups of children. The table also shows the χ^2 values between the proportions of above- and below-median instances for each structure and the levels at which they may be accepted as significant. Only three probability values are used, namely, 0·01, 0·02, and 0·05; and χ^2 values are corrected for continuity where the number of instances in any cell of the fourfold table is less than 50.

Consider, first, the results for the 16-year-olds. Structure II, which requires converting, is more difficult than I which is isotropic to start with. III, which requires reordering, is more difficult than I. III is also more difficult than II, which suggests that reordering requires more time for its performance than converting. Lastly, IV is the most difficult of all. These findings accord with what was said above concerning the reorganizings presumed to occur in the solving of three-term series problems.

Table 3

11-year-olds ($N=64$)					16-year-olds ($N=32$)				
	No. of solutions					No. of solutions			
Premise structure	Below median	Above median	χ^2	P	Premise structure	Below median	Above median	χ^2	P
II	157	99			I	91	37		
			3·581	—				3·855	0·05
I	133	123			II	75	53		
			3·129	—				6·252	0·02
III	113	143			III	55	73		
			0·511	—				6·888	0·01
IV	105	151			IV	34	94		

Compared with the 16-year-olds, the results for the 11-year-olds have three noteworthy features. (1) There is a reversal of the relative difficulty of I and II. This is strong evidence for the operation of a positive atmosphere effect in the younger children's solving of II. (2) This reversal apart, the trends (if not the magnitudes) of the inter-structure differences are the same as for the older subjects. As before, IV is the most difficult: it is not more difficult than III to a statistically significant extent, but it is significantly more difficult than I ($\chi^2=6{\cdot}157$) and than II. Confirmation of the difference between IV and I comes from results obtained by Burt in using his test no. 12. The two versions of this test were as follows. 'Edith is fairer than Olive: but she is darker than Lily. Who is darker—Olive or Lily?' 'Lily is fairer than Edith: Edith is fairer than Olive. Who is the fairest—Lily or Olive?' The first version conformed to structure IV and was solved by 46% of Burt's 8-year-olds: the second version conformed to structure I and was solved by 72% of the 8-year-olds. Another result of Burt's which agrees with the present findings occurs in connexion with his test no. 16. This test was as follows. 'Three boys are sitting in a row: Harry is to the left of Willie: George is to the left of Harry. Which boy is in the middle?' This problem was of type III and was solved by almost exactly 50% of the 9-year-olds studied. When the premises were reordered to make the problem of structure I, it was now solved by 61% of 9-year-olds. This difference of 11% between I and III is small and probably falls short of statistical significance as does the corresponding difference in the data of the present experiment. But in both Burt's case and the writer's, the difference between I and III is in the same direction; this suggests that III is indeed slightly more difficult than I for the younger children. The inter-structure differences, then, may be taken as having the same directions as those shown by the 16-year-olds (excepting the difference between I and II). But these differences are in every case smaller, and in most cases not statistically significant—a finding which is all the more striking when it is remembered that the number of younger subjects was twice that of the older. This raises the third feature of the 11-year-olds' results. (3) The amount of inter-structure difference is much less than for the 16-year-olds. The same inter-structure pattern is present, but is, as it were, blurred. This 'blurring' of differences reflects the fairly general finding that the judgments of younger children are more variable than those of older children both within the individual, and, the effects of specialized training apart, between individuals (Hunter, 1954). But more specifically, this 'blurring' indicates that the problem solving activities of the younger children are less influenced by variations in the problem structure and that these children are, as will be argued below, less appreciative of the structural characteristics of the problem situation.

(iii) *Effect of question asked*

Each of the four premise structures is followed by one of two questions. Does this difference in question affect ease of solving? Examination of the data shows that it does so in two significant cases only, both occurring with the 16-year-olds. For these older subjects, III, 2 is easier than III, 1 at the 0·02 level ($\chi^2 = 6{\cdot}248$); and IV, 1 easier than IV, 2, at the 0·05 level ($\chi^2 = 4{\cdot}083$). Both differences seem amenable to an explanation which is derived in the first instance from the writer's own retrospecting, and which, considering the present data as a whole, seems the only plausible explanation. The writer noticed that reorganizing served to isolate or emphasize one particular term of the series. Consider the premise pair (III) '*B* is shorter than *A*:*C* is shorter than *B*'. In deriving a series from these premises, the writer was very much aware of finding that *C* interrupted the direction of the series and of bringing forward *C* as the shortest member of the series. This made it possible to answer straight away the question 'who is shortest?' whereas the question 'who is tallest?' required a further reconsidering of the two remaining terms to decide which was indeed the tallest. If this introspective evidence is generalized into a formal statement, it would be: the term which is contained only in that premise which has to be reorganized is isolated on the ground that it should come at one end or the other of the entire series. This generalization is consistent with the 16-year-olds requiring more time to solve III, 1 than III, 2. It is also consistent with the slight, but not significant ($\chi^2 = 2{\cdot}355$), tendency for II, 1 to be more difficult than II, 2; and also with IV, 2 being more difficult than IV, 1 (provided the reorganizing involved in IV is of the reversing-converting rather than of the converting-reordering type, i.e. provided it is the first premise which is converted). If this explanation holds, it furnishes some evidence for problems of structure IV being solved by a reversing-converting procedure. Additional evidence in favour of this method of solving comes from a comment made by Burt in connexion with his test no. 12. Burt observed that, when the children were afterwards 'explaining their reasons' for their answers, nearly every child who had answered correctly stated that he had converted the disparate (i.e. the first-encountered) premise. To return to the present study, it is noteworthy that, with the 11-year-olds, the question asked has the same qualitative effects on solving time as it does with the older children but, in the case of these younger children, these effects are (as the reader might, by now, expect) not statistically significant.

(iv) *Effect of relation used*

Previous investigators have found that problems with familiar terms tend to be easier than structurally identical problems with unfamiliar terms (Wilkins, 1928; Brownell & Stretch, 1931; both as cited by Johnson, 1955, pp. 244 and 249). Here, it was hoped to demonstrate the effect of differences in the relations embodied in problems of identical formal structure. Two relations were sought such that both would be reasonably familiar to the subjects while one would lead to greater difficulty than the other in reorganizing the problem situation. There are, at the least, two respects in which a comparison involving height might be easier than one involving happiness. First, differences in height can be visualized more readily; and most children seem to be visualizers. Secondly, differences in height involve only one dimension or aspect while differences in happiness are multi-dimensional, i.e. one person may be happier than another in a variety of different ways. *A* may be happier than *B* because of better health and *B* may be happier than *C* because of a clearer conscience; and the fact that manifold aspects have to be considered before deciding that *A* is therefore happier than *C* would tend to hold up the thinker for a longer time than the simple comparison of height. In short, while there are doubtless many complex and obscure contingencies in the environment and in the organism which go to make one attribute more abstract than another, it may be said that happiness is a more abstract attribute than height. From these considerations, the first expectation is that, in the present experiment, the eight problems involving the 'taller-shorter' relation will be more rapidly solved than the eight structurally identical problems involving the 'happier-sadder' relation. A second expectation concerns differences between the two age groups. From a developmental point of view, it might be supposed that the same contingencies (whatever their number and nature) which make for the relative abstractness of an attribute would also retard the individual's acquisition of this attribute as a means of categorizing and ordering the events in his surrounding world. Accordingly, the second expectation is that, provided the appropriate age range is being studied, the younger children will find it relatively more difficult than will the older to handle problems involving the more abstract attribute. Results confirm both of the above expectations. The first 32 11-year-olds studied solve the 'taller-shorter' problems in less than the median time in 145·06 (out of 256) instances while the 'happier-sadder' problems are similarly solved in only 110·93 instances. The difference between these proportions is highly significant ($\chi^2 = 9{\cdot}031$). With the 16-year-olds, the 'taller-shorter' problems are easier to a quite insignificant extent ($\chi^2 = 0{\cdot}633$). Thus, the younger subjects find the relatively abstract 'happier-sadder' problems more difficult than the structurally identical but relatively concrete 'taller-shorter' problems while the older subjects do not.

With the study of a further group of 32 11-year-olds, the opportunity was taken to contrast the 'taller-shorter' relation with the 'warmer-colder' relation. Problems involving this 'warmer-colder' relation are solved in less than the median time in 118·31 (out of 256) instances while the 'taller-shorter' problems are similarly solved in 138·95 instances. Although the difference between these two proportions is not negligible, it is not quite statistically significant ($\chi^2 = 3{\cdot}446$). It would seem that reasoning about entities which differ in degrees of warmth is, for the 11-year-olds, intermediate in difficulty between reasoning about differences in happiness and differences in height.

(v) *Effect of atmosphere*

The only direct evidence for the effect of atmosphere is that the 11-year-olds find II easier than I. The results do not reveal whether or not atmosphere operates in the case of the 16-year-olds; but if it does, it is insufficient to make II easier than the logically more straightforward I. The most parsimonious conclusion is that, with regard to II, a positive atmosphere operates in the case of the 11-year-olds but not in the case of the older children. Another conclusion, equally tentative, might be drawn with regard to the results of the 11-year-olds. It is that little or no atmosphere operates in IV. This conclusion is based on two rather slender lines of evidence. First, in IV, 'inclusiveness' would operate to suggest one specific and incorrect answer: also, as argued in the Introduction, the only atmosphere likely to operate in IV is that which derives from 'inclusiveness'. It follows that the most frequently given wrong answer ought to be that suggested by 'inclusiveness'. But in fact, each of the two possible wrong answers occurs equally often—suggesting that these errors arise from confusion rather than from any systematic source. Secondly, if any negative atmosphere operated in IV, it might be expected that this structure would give rise to much more difficulty than it does in fact. Thus, what evidence there is hints at the operation of little or no negative atmosphere in IV. Now if this is so, it suggests yet another conclusion. If 'inclusiveness' does not operate in IV, it probably does not operate in II either; so the positive atmosphere of II would derive largely if not entirely from 'direct statement'. This is not to suggest that 'inclusiveness' never operates to facilitate and retard the solving of three-term problems; 'inclusiveness' might well have operated had, say, the subjects been a year or two younger.

IV. CONCLUSIONS

This study began by analysing the class of three-term series problems into eight different types or logical structures. Factors likely to affect the solving of problems of each structure were then discussed in the light of Burt's pioneer investigation. One set of factors involved an atmosphere effect in which the global impression of the problem suggests one particular answer which may or may not be correct. The second set of factors involved analytical judgments such as would entail processes of reorganizing, namely, reordering, converting, and reversing: an additional process, that of isolating, was added later as a by-product of the above reorganizing processes. Just as Bruner, Goodnow & Austin (1956) used reception and selection activities to construct 'ideal strategies' for the attainment of concepts so, in the present study, these reorganizing activities were used to construct the optimally efficient procedure for solving each problem of a given structure, that is, the minimal sequence of transformations which the given statement must undergo to yield an isotropic series and, hence, the solution. A variety of three-term series problems were then presented to groups of 11- and 16-year-olds on the assumption that the time taken to solve each problem would be a function of the factors mentioned above.

The main findings are summarized in the abstract at the beginning of this paper. But to these the following must be added. First, the older children require less absolute time to solve the problems. Secondly, the older children find converting easier than reordering. Thirdly, there is some evidence that problems of structure IV are solved by reversing-converting rather than by converting-reordering. Fourthly, the atmosphere effect shown by the younger children appears to affect only the solving of problems of structure II and to derive from 'direct statement'. Lastly, as far as premise structure is concerned, the order of increasing difficulty is I, II, III, and IV for the 16-year-olds and II, I, III, and IV for the 11-year-olds.

The several findings of this study are self-consistent and seem to exemplify two developmental trends. The first is that increasing age brings increasing appreciation of the structural characteristics of series. The second is that increasing age brings increasing skill in handling progressively more abstract relations, that is, relations which are progressively more complex and more remote from the perceptual-motor level at which, as Piaget has shown, the performance of seriating is first accomplished by the developing child. The first of these two trends must now be discussed in more detail.

'When I try to work out what is involved in the step sequences of my experiments... there seems to me to be two outstanding constituents in the thought process.' One of these is 'the appreciation of the structural characters of the system within which thinking is being exercised'. It is in these words that Bartlett (1956, p. 449) echoes, to some extent, the conclusions of Wertheimer (1945) and alludes to a feature which would seem to be of the first importance in directed thinking. What is meant by 'appreciation of structural characteristics' may be illustrated by citing one of the present writer's as yet uncompleted studies. Students are presented with a succession of simple five-premise sorites such as the following. 'All boys with red hair have grey eyes: all tall boys have red hair: all thin boys work hard: all boys with grey eyes are thin: all boys who play football are tall.' The student is asked to draw that conclusion which follows when each and every one of the above statements is taken into account and to 'think aloud' while doing so. When the student first tackles such a problem, his activity is haphazard. He may combine a couple of premises, draw a part conclusion, leave it, combine a further pair of premises, draw a second part conclusion, and try to see if this can be combined with the first part conclusion. In short, he solves the problem by a sequence of steps which is erratic and subject to many chance factors. But after solving a few more problems of this type, his performance is characteristically transformed. He largely ignores the order in which the premises are presented: he reads through the statements in search of that one of the two terms ('retinends') which are not repeated and which he will take as his starting point: and from this starting point, he considers each premise in such a sequence that the terms form a consecutive chain with identical terms ('eliminands') juxtaposed. Not only does he solve the problem in a shorter time but he does so in a more systematic way for, once he has selected his initial 'retinend', the sequence of steps taken is fully predictable. At this stage, he has, in his own words, 'got the hang of the problem'; in Bartlett's words, he appreciates the problem's structural characteristics. The difference between his initial and later performance is comparable to the difference between the execution of a well-articulated and practiced perceptual-motor skill such as typewriting and the performance of that same skill by a less practised person with his lack of assurance and his hunt-and-peck methods. It is not a rigid sequence of part-activities which has been learned but a co-ordinated and flexible system of activities from which the person can readily select those appropriate to deal with particular circumstances.

To return to the three-term series problems, the difference between the younger and older children is that the latter approximate more closely to this ideal appreciation of structure, that they behave, in a manner of speaking, more logically by more nearly following the optimal procedure for solving each problem. They can more readily treat the given premises as components of a series and more readily perform whatever reorganizings may be necessary in order to derive the series most expeditiously from them: they more easily recognize, and during the course of the experiment come to recognize, each problem as representative of a class of problems; and they know better just how to proceed in progressing towards the solution. The older children, as contrasted with the younger, have, the writer would suggest, a better appreciation of the structural characteristics of series problems. This better appreciation manifests itself in their solving the problems more rapidly, in their behaving in closer accord with the logical demands of the problem structure, in their resisting the suggestiveness of diffuse global impressions, and in their increasing to a greater extent their problem-solving efficiency over successive problems even though the pattern of reorganizings varies from problem to problem. The younger subjects arrive at the solution by less economic and predictable reorganizings with the result that their solving-time is both longer and less affected by variations in the logical structure: and being less able to analyse out the essential structure of the problems, they are more influenced by such irrelevances as global impression and less able to react to similarities between successive problems.

The writer wishes to thank Mr Iain Murray for conducting the experiment with the 16-year-olds. Thanks are due to Mr Mowat, Miss Brown, Miss Primrose, and Mr Spears of Moray House School, Edinburgh, and to Mr Murray of Portree Senior Secondary School for their kindness in making both subjects and testing facilities available. The writer also wishes to thank Sir Cyril Burt for much detailed and highly constructive comment on earlier versions of this paper.

REFERENCES

Bartlett, F. C. (1956). Some experiments about thinking. *Proc. Roy. Soc.* B, **145**, 443–51.

Bruner, J. S., Goodnow, J. J. & Austin, G. A. (1956). *A Study of Thinking.* London: Chapman and Hall.

Burt, C. (1919). The development of reasoning in school children. *J. Exp. Ped.* **5**, 68–77 and 121–7.

Carroll, L. (1895). *Symbolic Logic.* As selectively reproduced in *The Complete Works of Lewis Carroll.* London: Nonesuch. 1939.

Hunter, I. M. L. (1954). Tactile-kinaesthetic perception of straightness in blind and sighted humans. *Quart. J. Exp. Psychol.* **6**, 149–54.

Hunter, I. M. L. (1957). Note on an atmosphere effect in adult reasoning. *Quart. J. Exp. Psychol.* **9**, 175–6.

James, W. (1891). *The Principles of Psychology.* Vol. II. London: Macmillan.

Johnson, D. M. (1955). *The Psychology of Thought and Judgment.* New York: Harper.

Piaget, J. (1921). Une forme verbal de la comparaison chez l'enfant. *Arch. Psychol.* **18**, 141–72.

Piaget, J. (1928). *Judgment and Reasoning in the Child.* London: Kegan Paul.

Piaget, J. (1952). *The Child's Conception of Number.* London: Routledge and Kegan Paul.

Pinard, A., Barbeau, G. L., Laurendeau, M. & Parant, C. (1954). *Tests Différentiels d'Intelligence.* Université de Montréal.

Stebbing, L. S. (1933). *A Modern Introduction to Logic.* London: Methuen.

Wertheimer, M. (1945). *Productive Thinking.* New York: Harper.

Woodworth, R. S., & Schlosberg, H. (1954). *Experimental Psychology.* London: Methuen.

Woodworth, R. S. & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *J. Exp. Psychol.* **18**, 451–60.

(*Manuscript received* 27 *October* 1956)

THE INFLUENCE OF EXTRANEOUS STIMULATION ON THE REACTION TIME OF SCHIZOPHRENICS

By J. TIZARD and P. H. VENABLES

Medical Research Council, Social Psychiatry Research Unit,
Institute of Psychiatry, London

In a first experiment 74 chronic schizophrenics made 50 consecutive responses to the presentation of a stimulus light on each of two occasions. In an experimental run a 70 db white noise sounded between responses 21–30; in the control run there was no extraneous stimulation. Patients rated as 'withdrawn' showed an increase in speed of reaction in the presence of the noise, while there was no evidence to show that the extraneous stimulus affected those patients rated as 'sociable'. There was a general tendency for all schizophrenics regardless of classification to show a greater fall off of performance over time than normals. In a second experiment 30 sociable and 30 withdrawn patients responded to an auditory stimulus, in two different conditions of illumination. RT's of withdrawn schizophrenics were significantly faster when they worked in a brightly lit room than when they worked in a dimly lit room. Sociable patients showed a slight tendency to work faster in dim illumination. The findings are discussed.

I. Introduction

In a previous study (Venables & Tizard, 1956) it has been shown that the performance of non-paranoid, chronic schizophrenics on a continuous task differs from that of patients suffering from psychotic depression in that the schizophrenics show a more rapid decrement in performance, and greater reminiscence effects following a rest pause. In contrast to the schizophrenics, depressed patients show a typical 'reversal of the fatigue effect'. This finding confirmed earlier work by Kraepelin (1913) and by Hoch (1901) using different types of task, and related the phenomena to a single task. In the present experiments both paranoid and non-paranoid schizophrenics were tested in order to extend the scope of findings previously obtained, and to compare the two groups in their response to additional stimulation. Preliminary experimentation had indicated that those subjects who could clinically be thought of as withdrawn responded in a different way from those who who appeared more extraverted. To put this clinical impression on a firmer basis a scale which was intended to measure withdrawal versus sociability was constructed.

The present experiments were initially based on the following assumptions: if, as suggested in our previous study, the reason that schizophrenics fall off rapidly in performance is because stimulus-responses connexions lead to an abnormally rapid build-up of internal inhibition (or $\dot{I}_R$ in Hullian terms, for at present it is not proposed to distinguish between the two), then a 'distracting' stimulus should have the effect of dissipating some of this inhibition, and should lead to an increment in performance. Such an effect has in fact been found by Mackworth (1950) using normal subjects, and the first experiment to be reported here showed that 'withdrawn' schizophrenics speeded up under the influence of a distracting stimulus. This experiment did not however show that the mechanism responsible for the phenomenon was that of disinhibition; and a second experiment, designed to minimize the possibility of facilitation occurring through disinhibition was therefore carried out to see if the effect could still be demonstrated. An alternative hypothesis, that the effects were brought about by the 'tonic' effect of non-specific sensory stimulation on the cerebral cortex, was also proposed.

The aims of the experiments can be summarized as follows. (1) To compare the work decrement of paranoid and non-paranoid schizophrenics, and withdrawn as against sociable patients, on the reaction time task. (2) To test the theory that an extraneous stimulus, introduced into a reaction time series, will have a facilitating effect on the performance of schizophrenics. (3) To see whether schizophrenics rated as sociable (on the basis of a rating of social withdrawal by charge nurses or ward sisters) react differently under conditions of extraneous stimulation from those rated as withdrawn.

II. First experiment

(i) *Procedure*

Subjects were tested individually under standard conditions. They were required to respond as quickly as possible to the appearance of a visual stimulus by pressing down a toggle switch. A toggle switch rather than the conventional morse key was used, as its action was more clearly understood by seriously ill patients, and as the absolute speed of RT was not under consideration. The visual stimulus consisted of a ground glass screen 0·75 in. in diameter, illuminated from the rear, set in a 2 ft. by 2 ft. matt black panel. The intensity of the stimulus was 135 ft. candles, and the background illumination reaching the subject's eye was 1·7 ft. candles. A bell served as a warning signal preceding the appearance of the visual stimulus by 8·0 sec. The stimulus was presented for 3 sec. if no response was made, and the cycle of bell and light was repeated once every 15 sec. Each subject responded to 50 stimuli running consecutively; in the experimental series the extraneous stimulus sounded continuously during responses 21–30. The subjects were told that they might hear some 'static'—'like you get on the radio sometimes'—during the experiment, owing to 'interference', and this explanation proved sufficient for them to accept the situation naturally when the noise occurred.

Half the subjects performed first under control conditions, that is, they made 50 consecutive responses with no extraneous stimulus; the other half were tested first under experimental conditions, that is, where the additional extraneous stimulus sounded between responses 21–30. On a second occasion of testing the subjects worked under the condition which they had not previously experienced.

The extraneous stimulus was white noise produced by amplifying the discharge of a neon tube by means of a circuit having linear response from 20 to 18,000 c.p.s., and was reproduced by means of a single 12 in. diameter loud-speaker set in a 4 ft. by 4 ft. baffle situated 5 ft. from the subject's head. The intensity of the sound was adjusted to be 70 db as measured by means of a noise level meter. The intensity of both stimulus light and sound stimulus was kept constant by being supplied from a regulated voltage source.

Reaction times were measured by means of a cold cathode counter tube timer to the nearest millisecond.

(ii) *Subjects*

Forty-two male and 32 female subjects of whom 46 were classified as non-paranoid schizophrenics and 28 as paranoid schizophrenics were tested. All were less than 56 years of age and had been in hospital more than 1 year. The mean age of the male patients was 40·85 s.d. 8·94 years, and of the female patients 45·5 s.d. 9·19 years. No patient had undergone leucotomy, but some patients were under treatment by chlorpromazine, methopromazine, and similar drugs at the time of the investigation. (The effect of these treatments upon performance is described below.) Six male and 6 female nurses acted as normal controls. Their mean age was 26·16 s.d. 4·19 years.

Each patient was rated by two charge nurses or sisters on a scale constructed to measure sociability-withdrawal. There were in this scale 6 items, 5 of which were taken from the M.S.R.P.P. Scale developed by Lorr, Singer & Zobel (1951). (Items nos. 8, 52, 56, 57 and 58.) An additional item was included to give a measure of the number of friends the patient had in the ward. The average inter-rate reliability was $r=0{\cdot}765$ for an N of 74. Reliabilities of 0·854 and 0·675 were obtained respectively for charge nurses rating male patients, and ward sisters rating female patients. Independence of the scale from age is shown by a correlation of ±0·064 between the subject's score and age in years.

(iii) *Results*

For the purpose of the analysis, the 50 responses made by each subject were considered in 5 groups of 10 each, and further work was carried out on the mean values of each group of 10 responses. As the distribution of RT's and particularly of changes in RT under extraneous stimulation were not normally distributed, non-parametric statistics were employed. Because in this instance mean values of RT tended to obscure the picture owing to the abnormally large responses by some subjects, the results were analysed on the basis of change in direction of responses without taking note of their magnitude.

To examine the trends in performance of the schizophrenic subjects under the control conditions the performance of each subject during the control run was summarized by a linear regression line, and the slope constant examined. In 55 out of 74 subjects the slope was positive, indicating that RT became slower throughout; this as it stands is significant at the 0·01 level by the Sign Test (Dixon & Mood, 1946). In the case of the six normal controls, the slope was negative in five cases. Although the application of a χ^2 test in the case of such small numbers in two cells must be considered with caution, the result of 6·245 after applying Yates' correction is probably sufficient to indicate a significant association between the increase in RT and schizophrenia, as the size of the χ^2 gives probability value beyond the 0·02 level of significance.

Further examination, to determine whether the increase in RT during the control runs is associated more with paranoid than with non-paranoid schizophrenia, or is found more in withdrawn than in sociable schizophrenics, results in insignificant χ^2 of 0·016 and 0·621 respectively. The age difference of subjects showing positive or negative slope constants is not significant ($P > 0{\cdot}05$). This lack of association with age confirms the previous finding (Venables & Tizard, 1956), where there was no significant association between decrement in performance on a repetitive task and age.

The test of the hypothesis that schizophrenic performance can be speeded up by an extraneous stimulus was made by examining the change in RT between the means of responses 11–20 and those of responses 21–30. The number of subjects whose mean RT's were greater during responses 21–30 than during responses 11–20, that is, who were slower while the extraneous stimulus was sounding, was 35, whereas the number of subjects who were faster while the extraneous stimulus was sounding was 39. The difference is not, of course, significant, and the second hypothesis—that an extraneous stimulus would facilitate the performance of schizophrenics—must be regarded as untenable.

The data were then examined to compare the performance of paranoid and non-paranoid subjects, and sociable and withdrawn subjects. Of the non-paranoid subjects, 29 were faster while the extraneous stimulus was sounding, while 17 were slower, and of the paranoid subjects 10 were faster, and 18 were slower. This association between diagnosis and change in response speed is significant as shown by a χ^2 of 5·21, $P < 0{\cdot}02$.

To see whether 'social withdrawal' was a relevant factor in influencing change in response under an extraneous stimulus, patients were divided into two groups depending on whether they fell above or below the half-way point on the sociability rating. Of the withdrawn subjects, 23 were faster during the time the auditory stimulus was sounding, while 8 were slower. Of the sociable group 27 were slower, and 16 were faster. These figures gave a χ^2 of 9·89, which is significant at the 0·01 level. The difference between the mean ages of subjects increasing and decreasing in RT under extraneous stimulation is not significant ($P > 0{\cdot}05$).

In a normal control group who, for administrative reasons, were tested only once, two subjects increased and two decreased in RT while the auditory stimulus was sounding, while two remained the same. It is apparent that the auditory stimulus had little effect on the normal subjects.

As a check it is now desirable to examine the performance of the schizophrenic subjects under control conditions. The number of subjects showing a decrease in reaction latency from responses 11–20 to 21–30 is 28. The number showing an increase in latency is 46. If these subjects are now divided as before into paranoid and non-paranoid groups, among the paranoids 12 show a decrease in reaction latency, and 16 an increase; and of the non-paranoids 16 show a decrease in latency, and 30 an increase. That the change in latency under control conditions is not relevant to diagnosis is shown by an insignificant χ^2 of 0·504. A similar division into withdrawn and sociable groups shows by an insignificant χ^2 of 0·228 that there is no relationship here either.

The analysis so far has shown that both patients diagnosed as non-paranoid, and those rated as withdrawn, tend to decrease in RT (i.e. they were faster), while the extraneous stimulus was sounding; patients diagnosed as paranoid or rated as sociable increased in RT. However, it was noted that during the control run there was a tendency for all groups to increase in RT throughout the testing session. The data were therefore re-examined to see if the trend in performance during the extraneous stimulation differed from the trend shown by the same patients during the control run. A ranking technique was employed to make the comparisons.

If we consider the five periods of 10 responses, and assign a rank value to each on the basis of magnitude of RT we can compare the ranks obtained by each subject under the experimental conditions with the ranks of the same subjects under control conditions. The mean values of the ranks actually obtained are shown in Table 1.

Table 1. *Mean ranks of response time for five periods of* 10 *responses under control versus experimental conditions* (*Rank* 1—*fastest response*)

	Period					
	1	2	3	4	5	Total
Paranoid $N = 28$						
Experimental	2·95	2·64	2·96	3·23	3·22	15·00
Control	2·96	2·68	2·66	2·57	3·13	15·00
Non-paranoid $N = 46$						
Experimental	2·61	2·98	2·26	3·66	3·49	15·00
Control	2·34	2·63	2·92	3·51	3·60	15·00
Sociable $N = 44$						
Experimental	2·57	2·61	2·92	3·61	3·28	15·00
Control	2·87	2·66	2·66	3·53	3·27	15·00
Withdrawn $N = 30$						
Experimental	2·98	3·20	1·95	3·33	3·54	15·00
Control	2·13	2·63	3·07	3·53	3·64	15·00

It is seen from Table 1 that with the sociable and paranoid subjects the difference between the ranks obtained during Period 3 under control and experimental conditions is slight, being 0·26 and 0·30 respectively. These differences are not significant ($t = 0{\cdot}870$ for the sociable group and 0·816 for the paranoid group). We must conclude therefore that any increase in RT is not due to the extraneous stimulation, but is part of the general pattern of response.

In the case of the withdrawn and non-paranoid subjects there is a greater difference between the ranks obtained during Period 3 under the two conditions ($C3-E3=1{\cdot}11$ and $0{\cdot}66$ respectively). These differences are significant in both cases; ($t=3{\cdot}42$, $P<0{\cdot}01$, and $t=2{\cdot}26$, $P<0{\cdot}05$ respectively). We can therefore conclude that the effect of extraneous stimulation on these two groups of subjects is that of facilitation.

Owing to the small number of subjects (six), who were in the withdrawn paranoid group, further analysis of the data to investigate the interaction between the two factors of diagnostic category and social withdrawal yielded no significant results.

The responses of patients being treated with chlorpromazine and similar drugs were then examined. Of 23 such patients, 18 were rated as sociable, and only 5 as withdrawn. With these subjects the association between decrease in RT while the stimulus was sounding and social withdrawal was weakened. χ^2 is reduced to 2·48 which does not reach an acceptable level of significance, probably because of the small number of withdrawn schizophrenics in this group.

(iv) *Discussion*

The findings are as follows. Under control conditions performance of schizophrenics in this as in other experiments shows a significant 'fatigue effect', the extent of which is independent of diagnosis, age or 'social withdrawal'.

The effect of an extraneous stimulus was to facilitate the performance of non-paranoid and withdrawn schizophrenics. There was no evidence to show that a distracting stimulus affected the performance of sociable or paranoid patients.

The findings as they stand give partial confirmation of the prediction from the hypothesis that the RT's of schizophrenics can be speeded up through disinhibition in so far as one group of schizophrenics, whether by disinhibition or not, showed facilitation in their RT performance in the presence of a loud auditory stimulus. An alternative and perhaps more likely hypothesis is that in chronic, withdrawn schizophrenics the central excitatory state is lowered in some way. An increase in sensory input might be expected to have non-specific facilitatory effects on cortical and subcortical activities. Electrophysiological work summarized by Lindsley (1956) suggests that such effects can be found in animals, and work associated chiefly with the name of G. L. Freeman (e.g. 1948) has indicated that in normal human subjects excitability to various specific stimuli may be increased by an increase in background stimulation.

A second experiment was therefore carried out under conditions which reduced the possibility of facilitation occurring through disinhibition to a minimum. In this experiment RT's to an auditory stimulus were measured under two conditions of background illumination to see whether the additional non-specific sensory input coming from a brightly lit room would have facilitatory effects.

III. SECOND EXPERIMENT

(i) *Procedure*

Forty-eight male and 12 female schizophrenics all less than 55 years old and whose mean age was 42·5 years S.D. 9·26 years, served as subjects. Thirty subjects were rated as 'sociable' using the procedure described earlier, while thirty were rated as 'withdrawn'. Twenty-one patients were being treated by chlorpromazine at the time of the experiment.

Each subject made 20 responses to a 1000 c.p.s. tone whose intensity was 60 db above threshold, presented through a loud-speaker under bright (18·0 ft. candles) illumination and 20 responses to a similar tone in a dim (0·3 ft. candles) illumination. Half the subjects worked first in bright and then in dim illumination, and half in dim and then bright illumination. The allocation of subjects was made randomly. The subject was allowed 3 min. in each condition of illumination to become accustomed to the situation before being tested. RT's were measured to the nearest millisecond by means of an electronic timer.

(ii) *Results*

Preliminary analysis showed that the order of illumination in which the subjects performed was not a significant variable; ($\chi^2 = 0{\cdot}0026$, $P > 0{\cdot}90$). The first part of Table 2 shows the number of subjects, divided on the basis of the sociability-withdrawal rating into 'withdrawn' or 'sociable' groups, who responded more quickly under bright or dim illumination. The association between rating and response pattern is highly significant ($\chi^2 = 11{\cdot}84$, $P < 0{\cdot}001$). Of the sociable patients 19 were faster in dim illumination while 11 were faster in bright; and of the withdrawn patients nine were faster in dim illumination while 21 were faster in bright. As in the first experiment there is no statistically acceptable evidence that sociable patients are faster in one condition than in the other ($P > 0{\cdot}25$ using the Sign Test). The probability that the results obtained with the withdrawn patients is due to chance is however less than 0·05, using the Sign Test.

Table 2. *Numbers of subjects giving faster RT's in dim or bright lights*

	Total group		Paranoid subjects		Subjects under drugs	
	Sociable	Withdrawn	Sociable	Withdrawn	Sociable	Withdrawn
Faster in dim light	19	9	3	0	8	0
Faster in bright light	11	21	1	3	5	8
χ^2	11·84		—		—	
P	0·001		—		—	

All but 7 of the patients were diagnosed as suffering from other than paranoid schizophrenia. Separate data from the seven paranoid subjects are shown in the middle part of Table 2, where it is seen that they behave according to the general pattern. The same tendency is shown by the 21 patients under drug treatment at the time of the experiment, whose data are given on the right hand side of Table 2.

(iii) *Discussion*

The results show that in an experimental situation in which the possibility of disinhibition as a mechanism for facilitation had been minimized, there was again a significant tendency for withdrawn patients to react more quickly when an extraneous stimulus was present. The alternative hypothesis, that an increase in sensory input has a non-specific facilitatory effect in withdrawn patients thus secures support.

This work was carried out in Springfield Hospital, London, by kind permission of the Medical Superintendent, Dr H. C. Beccle.

REFERENCES

DIXON, W. J. & MOOD, A. M. (1946). The statistical sign test. *J. Amer. Statist. Ass.* **41**, 557–66.

FREEMAN, G. L. (1948). *The Energetics of Human Behavior.* New York: Cornell University Press.

HOCH, A. (1901). On certain studies with the ergograph. *J. Nerv. Ment. Dis.* **28**, 620–8.

KRAEPELIN, E. (1913). *Psychiatrie* (vol. 3). Leipzig: Barth.

LINDSLEY, D. B. (1956). Physiological Psychology. *Ann. Rev. Psychol.* **7**, 323–48.

LORR, M., SINGER, M. & ZOBEL, H. (1951). Development of a record for the description of psychiatric patients. *Psychol. Serv. Cent. J.* **3**, no. 3.

MACKWORTH, N. H. (1950). *Researches on the Measurement of Human Performance.* M.R.C. Special Report, no. 268. London: H.M.S.O.

VENABLES, P. H. & TIZARD, J. (1956). The performance of functional psychotics on a repetitive task. *J. Abnorm. (Soc.) Psychol.* **53**, 23–6.

(*Manuscript received* 11 *October* 1956)

THE BOUNDARIES BETWEEN THE PERSON AND THE WORLD

A NOTE ON THE PAPER BY GARDNER MURPHY (*Brit. J. Psychol.* 1956, **47**, 88–94)

BY LEON LITWINSKI, *Brussels*

Like Murphy I appreciate the field theories of Lewin and Angyal to which he refers in his paper (see, Litwinski, 1951, 1956). My formulations, however, differ in some respects from Murphy's. In contrast to the widespread opinion that person and world can be sharply distinguished, Murphy emphasizes that one often melts into the other. In fact I made this same point in a paper presented to the Eleventh International Congress of Philosophy. There I quoted Ward (1951) who speaks of a pilot's feeling of one-ness with his machine. But, in my opinion, such cases must not be generalized too much. They belong rather to animistic states of mind which prevail in certain periods of development or in certain circumstances. They also belong to selfless experiences or impersonal states. They should be called identifications rather than fusions. Therefore, the notion that we tend to melt into our environment seems to be rather a metaphorical way of speaking. Like all analogies, they are safe if not pushed too far. The scientific statements require some rigorousness in the use of terms.

Under the command of love, for instance, a conjugal union may become so intimate that it may be called '*égoisme à deux*'. Are we to conclude that the distinction between the two members of such union vanishes and that it is permissible to speak of osmotic fusion of two human beings? It seems that we must not go too far in formulating our theories in this way. And we must recognize a figurative mode of speech for what it is. 'If two persons of the same mind (*natura*) come together,' said Spinoza 'they form an individual twice as strong as either....Nothing, therefore, is more useful to man than man.' But the contrary may be argued: if two men are starving and there is only one loaf of bread, it is soon apparent that nothing is more detrimental to a man than another man. The man who finds that the woman he loves has another lover, makes a similar sad reflexion. Hence, in theoretical considerations there is constant need for stress on rivalry, on states of crisis or drama in human intercourse, resulting from the identity or similarity of individual needs and the too often limited means of satisfying them. We have to distinguish what the man seeks to get, what he can get and what he has got. Plotinus emphasized that man gets what he can and keeps what he has got. (*Lucri quidquid est, id domum trahere oportet.*) The constant pressure of the outer and inner worlds tend to create for the individual a number of vital problems, of different magnitude, ranging from physical sufferings and discomforts to moral pains and uneasiness. So arises the search for cohesion, coexistence, adjustment, etc.

If it is true that, in the world of physics, man can conquer nature on condition that he obeys her laws, it is also true that, in social affairs, man can more freely integrate himself within a group if he respects the liberty of his fellow men. Here emerges the need to appreciate those limits that are functionally related to belongingness. Such a sense of limitation may be either underdeveloped or overdeveloped (another source of crisis extending from propensities to waste to those of avarice). By this sense of limitation however the individual may reveal wisdom or foresight. Recognition of limits may thus be therapeutic in education as well as in politics where stiffeners as well as softeners are required. The numerous material bonds, the various affectional or social ties are all relative in meaning. They may be strong or weak, lasting, ephemeral or transitory, developed or undeveloped. Their manifestations are observable not only in space but also in time. This is why, in my studies of personality, I stress the prospective character of bonds and ties between the subject and the 'object' considered in its extended sense, i.e. signifying not only material things.

To the fundamental question, are we one with the world or isolated from it, my reply is that we are neither one with the world nor separated from it. In varying degrees, the individual is prolonged in the outer world through his belongingness and he is projected in the future through his expectations. Thus, in my view, both the concept of belongingness and the notion of expectation are particularly important in the interpretation of the facts.

REFERENCES

LITWINSKI, L. (1951). Toward the reinstatement of the concept of self. *Brit. J. Psychol.* **42**, 246–9.

LITWINSKI, L. (1956). Belongingness as a unifying concept in personality investigation. *Acta Psychol.* **12**, 130–5.

WARD, T. H. G. (1951). The psychological relationship between man and aircraft. *Brit. J. Med. Psychol.* **24**, 283–90.

(*Manuscript received* 30 *July* 1956)

PUBLICATIONS RECENTLY RECEIVED

Secondary School Selection. Ed. by P. E. Vernon. London: Methuen. 1957. Pp. 216. 15*s*.

'Secondary School Selection' is the first of what, it is hoped, will be several volumes on topics of public interest, published under the auspices of the British Psychological Society. It is the report of a working party of fourteen psychologists set up by the Society to give an authoritative statement of psychological research and opinion on the subject. The authors discuss at length the psychological, statistical and technical aspects of selection, but aim nevertheless to show that its problems arise primarily from historical, administrative, political and social causes, which cannot be removed simply by changing or abolishing existing methods. The report is addressed to educational administrators and teachers as well as to psychologists, and ends with thirty-two recommendations or spurs to action which serve to remind us that contemporary psychology is a profession as well as a field of academic study. It is chock-a-block with facts, arguments and theories, in which the influence of Prof. Vernon, who has edited the text, is clearly to be seen. Few psychologists will agree with all the findings put forward, but no one familiar with the problem could fail to appreciate the thoroughness and objectivity with which this task has been carried out, and it has certainly succeeded in showing how large a contribution psychology has made to this problem of national importance.

This book should go a long way to dispel the notion that the eleven plus examination is strongly supported by all psychologists, most of whom, it asserts, dislike the system, and look to alternative types of school organization based on the concepts of allocation (blessed word) and guidance rather than competitive selection: psychologists would apparently abolish the selection system if there were a practicable and just alternative, and many are said to have gone so far (and more haven't) as to advocate that their profession should 'contract out' and refuse to have anything to do with procedures that have such harmful effect on children's emotional and educational development. The committee takes the view that it is for society and its educational leaders to decide *whether* modifications in the selection procedure are desirable. As social scientists, psychologists should not, it is thought, feel committed to attack or to support the present system. Their function is to investigate which features are psychologically beneficial or harmful to individuals, and to ensure that the procedures employed are the most appropriate and valid for a given set-up. This gives a clear opinion on the vital problem of the social responsibility of the educational psychologist, who should presumably proffer advice to education authorities but should not thereby regard himself as any more privileged to insist on its implementation than the man in the next polling booth. This seems wise in view of the social consequences attendant on grammar school entrance, but there must surely be limits to the extent to which some psychologists will wish to co-operate in selection procedures which they believe to have had psychological effects on the candidates. It is only too evident to psychologists who come into contact with parents in clinics, or in lecturing to parent-teacher associations, that the natural desire of parents to ensure the best opportunities for their children has helped to make the process of selection for secondary schools a source of widespread anxiety. The report errs perhaps in giving a rather piecemeal emphasis to clinical experience concerning the effect of the examination on the four-fifths of the population who fail it, even if the emotional effects on the children are probably less than the ill-feeling caused among the parents. It might well be worth exploring such propaganda techniques as the film and pamphlets, which would help to bring home to parents what is involved in educational guidance at all stages of school life, and a shortened version of the present report, written for the ordinary parent, might perform an even greater public service than the present document. The working party thinks that, in view of the disquiet whipped up by the press and by politicians, exaggerated statements should be corrected. In this respect, many psychologists prefer to ignore ignorant criticism and to believe that the slow accumulation of objective evidence of the sort presented in this report will eventually reach the public.

Educational administrators will find advice on numerous topics, including scaling techniques, the use of subjective data, age allowances, the expected number of grammar school misfits, the marking of essays, school records, teachers' estimates, the selection of children attending small schools, the delimitation of the border zone (50th to the 95th percentiles), various borderline procedures, individual and panel interviews, an account of the extremely thorough method used in the city of Lincoln, and the West Riding scheme for scaling against the junior schools' results in previous years, or retrospectively against the pupils' performance on tests given when they reach the secondary schools. The appendices might reasonably have included some material from the main text, as the discussion of such technicalities as multiple criteria and correction for attenuation and homogeneity are of little general interest, although everyone would be well advised to look at the appendices, if only to gain some insight into the complexities involved in selection. For example, Appendix B, on the constancy of the I.Q., develops the argument that the amount of variation will depend on the similarity of the tests employed, the time interval between testing and retesting, the chronological age of the subjects, the range of validity in the group, the level of ability, the short term reliability of the tests concerned, the standard deviations of the tests, the number of tests compared, and the range of the norms. The report emphasizes the very convincing evidence found in nearly every investigation of the superiority of the intelligence test as an overall predictor of later success, but fails to bring out clearly enough the additional advantage that fluctuations in I.Q., although larger than originally suspected, will usually be less than those in Arithmetic, English or any other selection variable. The important point is made that intelligence tests act much more as a class leveller than as a class perpetuator and that it would be more logical for left wing critics to welcome them than to decry them, since on both theoretical and experimental grounds it seems likely that in areas where the intelligence test has been abandoned fewer working-class children will gain grammar school places than in those areas where they still constitute part of the examination. Authorities who abolish intelligence tests are probably making a decision to withhold educational opportunity from working-class children.

Teachers will be interested to know that, in an important investigation, the scaled teachers' estimates gave correlations of the same order as those of the external examinations, and higher than those of the objective attainment tests, although the likely error of a teacher's unscaled estimate of the average ability of his class is twice that of the error of an estimate based on a mediocre intelligence test. 'Old type' examinations, when constructed by persons with experience of psychometric technique, can reach the same efficiency as the objective attainments test; and the essay, despite the unreliability inherent in its marking, does add appreciably to the validity of prediction, although no weight should be attached to a single marking of a single composition. The advice given about coaching is rather ambiguous. It is considered that coaching should be discouraged, but that if it cannot be controlled, adequate previous practice and explanation in all schools should be authorized. It is also admitted that the abilities of individual children are more variable than was earlier supposed, and that rigid streaming may have harmful educational and social consequences, although these are often exaggerated. And grammar school masters will, no doubt, agree that the G.C.E. is the nationally accepted standard for entry into higher forms, industry, commerce and the professional world, and so constitutes an important criterion of achievement.

The report will also serve as a text-book for students, and as a guide to lecturers in educational psychology. It includes sections dealing with the history of intelligence testing, theories of the nature of intelligence, the work of Hebb, the effects of environment and heredity on test scores, and the normality of I.Q. distributions. The excellent suggestion is made that the term 'general educability' should be substituted for 'general intelligence', on the ground that this term makes no claim that psychologists are measuring purely innate ability. On the other hand, it would be even more misleading to claim that we are measuring purely acquired attainment. And the work of Newman, Freeman and Holzinger is quoted merely to show that in very different environments twins with the same hereditary equipment might differ by as much as 20 points in I.Q., despite the fact that the main conclusion reached by these authors was that about 75% of the variance shown by tests of intelligence is attributable to innate or inheritable causes. Nor does this section on 'How Far Do Intelligence Tests Measure Inborn Ability?' refer to the extensive work of Burt in this field, which also points to the predominance of genetic factors. Research strongly suggests that in our present system differences in I.Q. scores are mainly determined by heredity. It is true, of course, that if our social environment were less homogeneous the environmental contribution would be increased, and that recent changes in the attitudes of parents, coaching, etc., may be altering the balance, but until contrary experimental results rather than hypotheses are advanced it would be more scientific to emphasize the innate component in intelligence test scores. Everyone would agree with the working party that psychologists should frankly acknowledge that a completely accurate classification is impossible at eleven years, but they should also add that perfect prediction of complex human behaviour at any age is impossible, and that psychological principles can no more be applied unerringly to individual children than the principles of physics can afford predictions of the exact path of a slate that falls off a roof. The important point is made that the best vindication for a system of selection is that it admits of fewer errors than would occur with any other method, and that we are now near the limits of predictive efficiency, which is a considerable technical achievement. Yet the claim on the dust cover that the procedure is accurate for some 90% of children may be disputed, since it assumes that the correlation of an imperfectly reliable selection procedure with a perfectly reliable scholastic criterion is close to plus 0·90. In fact, scholastic criteria are by no means perfectly reliable, and the effect of this correction for attenuation is to give a somewhat optimistic estimate of the efficiency of selection procedures, although the discrepancy is not very great. It must be remembered that (as stated in the text) a quarter of those selected, on present standards, do not deserve their place. And is it fair to state that in view of the high validities achieved we can hardly expect that 'tinkering with' selection procedure, e.g. including essays as well as subjective English tests or additional personality assessments, would produce any marked improvement? The addition of essays and personality estimates are scarcely 'tinkering', since they raise the reliability, mitigate backwash effects on the school curriculum, and may eventually replace the original tests without lowering the overall validity.

Statistical psychologists will enjoy the appendices. Appendix A, 'Follow Up Studies of the Validity of Selection', comprises tables from six representative researches. Appendix B, on the constancy of the I.Q., summarizes research results, and ends with a table showing the gains or losses in I.Q. points to be expected for identical tests over intervals of 1 month and 5 years respectively. Appendix C, on the reliability of scaling of primary school marks and estimates, criticizes some of McClelland's earlier conclusions, and Appendix D gives particulars of sigma and percentile scaling and discusses the methods used by various education authorities, including the very workable scheme adopted by the National Foundation for Educational Research.

A question raised in everyone's mind is whether this report is not bolting the stable door after the horse has gone, since secondary school selection seems likely to be dropped by several authorities. However, educational provision in this country is such that any immediate wholesale change seems improbable, and, in any case, the authors of this report believe that, if comprehensive schools become general, psychology will have at least as much to contribute to the techniques of diagnosis and guidance within such schools as it has to the more restricted aim of accurate selection.

F. W. WARBURTON

Theories of Personality. By CALVIN S. HALL and GARDNER LINDZEY. New York: Wiley. London: Chapman and Hall. 1957. Pp. xi+572. 52*s*.

Most teachers of personality theory would agree that there is no *one* useful survey of the existing important theories. This book is intended to make good the deficiency. In the author's words, 'It provides compact yet comprehensive summaries of the major contemporary theories of personality written at a level of difficulty that is appropriate for undergraduate or graduate instruction.' It is their hope that 'this volume will serve a function in the area of personality similar to that served by Hilgard's *Theories of Learning* in the area of learning'. Chapters of a modal length of 40 to 50 pages are devoted to the theoretical positions of the following: Freud; Jung; 'Social Psychological Theories' (Adler, Fromm, Horney, and Sullivan); Murray; Lewin; Allport; 'Organismic Theory' (Goldstein, Angyal, Maslow, and Lecky); Sheldon; 'Factor Theories' (Eysenck, and Cattell); 'Stimulus-Response Theory' (Dollard and Miller, Sears, and Mowrer); Rogers; and Murphy. There is also an introductory chapter on 'The Nature of Personality Theory' and a concluding chapter on 'Personality Theory in Perspective'. A praiseworthy attempt has been made to outline each theory in terms of a standard pattern of exposition: introduction and orientation (some of the biographical material here is most interesting); the structure of personality; dynamics of personality; development of personality; characteristic research and research methods; and finally, current status and evaluation. But this expository schema has not been rigidly adhered to when not suitable, e.g. with Sheldon's theory. Also, when a number of different writers are taken as representing one theoretical position, as are Adler, Fromm, Horney, and Sullivan, there is only one section on 'current status and evaluation'. This is sometimes to be regretted, as when it involves no separate evaluation on Eysenck's position. Where possible, each theorist has read the chapter devoted to his theory; this includes nine out of twelve chapters. To this the authors acknowledge their heavy indebtedness. Furthermore, parts of the book have been read and criticized by an impressive array of big names in personality theory, and the whole book has been seen by Allport.

This book will certainly prove invaluable to both teachers and students of personality. Here at last is a single book which expounds most of the important contemporary approaches to personality in a mature and reliable way. It should prove especially useful as a basic text for honours students doing courses in personality theory; in this respect at least, its avowed aim of becoming the Hilgard of personality theory does not seem immodest. At the same time its extensive references to both primary and secondary sources for each theory should enable the student to find his way into the literature. Criticism seems likely on two main counts. First, it is possible to disagree with the author's selection of theories; it might well be argued, for example, that the inclusion of a chapter on 'Organismic Theory' at the price of omitting such an historically important position as McDougall's (or indeed Shand's) and of devoting less than three pages to Mowrer is distinctly odd. There are those, too, who would question the wisdom of extended expositions of the views of Rogers and Murphy. But probably no two psychologists would agree completely about what to include and exclude. Secondly, it may be claimed that the authors' biases are from time to time too clearly evident, as for example in their somewhat lyrical account and appraisal of Murray's theory. On the whole, however, their expositions of the various theories are lucid and balanced; the chapters on Sheldon and on Lewin struck the reviewer as being especially good. The authors do not claim to present a full critique of each theory: their aim is essentially expository. Nevertheless, their sections on 'current status and evaluation' at the end of each chapter, though brief, are sometimes provocative; an instance is provided by their chapter on 'Factor Theories'.

Perhaps, the most remarkable and unusual quality of this book is the relative sophistication of its approach to the functions of a theory. This is especially evident in the first and final chapters, but also from time to time in the appraisals of the various theories (though here less often than one might wish). It is the authors' opinion that 'the field of personality would benefit enormously from an increased sophistication on the part of psychologists concerning the nature and function of theoretical formulation'. In the first chapter they outline a view of theory that is 'a relatively conventional summary of the thinking of methodologists or logicians of science'. Thus, a theory is looked upon as a set of conventions rather than as a fact of nature: the role of assumptions and of empirical (co-ordinating or operational) definitions is stressed, and the predictive functions of a theory are heavily emphasized. In the body of the text a constant criticism of theories is that they too often provide after-the-fact explanations and too seldom before-the-fact predictions. But the authors are not over preoccupied with *formal* considerations of theory adequacy, though they do emphasize that no existing theory measures up to these exacting standards. They draw a useful distinction between the way in which a theory may generate research by yielding, deductively, testable propositions about stable relationships between observables (classical hypothetico-deductive method), and the way in which, contingently, a theory may have heuristic value. For example, the authors believe that Lewin's theory, though formally inadequate, was heuristically very valuable, because it happened to inspire much research of generally accepted worth.

Another valuable feature of the book is the attempt made to compare the different theories with one another in terms of their relative positions concerning a dozen and a half important points of controversy on the content (rather than the form) of personality theory. Examples include the varying importance of purpose, of unconscious motivation, of reward, of contiguity, of heredity, and of the concept of the self, in the various theories. A table summarizing this 'Dimensional comparison of Theories of Personality' is given on p. 548. Interesting, too, as well as provocative are the authors' 'Some Reflections on Current Personality Theory' in the last chapter. What is particularly commendable here is the way they stress the value of empirical research that bears directly on existing theories: they castigate the current overconcern with theoretical originality, however trivial. Their rejection of the attractive, but facile, view that present theories of personality should be 'synthesized' is also refreshing.

In all, the authors are to be congratulated on producing a sophisticated, relatively well-balanced and reliable survey of a difficult, diffuse and highly controversial field of psychology. They will certainly receive the gratitude of many future students and not a few teachers.

C. W. PILKINGTON

Perspectives in Personality Theory. Ed. by HENRY P. DAVID and HELMUT VON BRACKEN. London: Tavistock. Pp. xiii + 435. 38*s*.

We are told in the last chapter that 'this book is a child of the Fourteenth International Congress of Psychology, which convened in Montreal in June 1954'. The 'child' proves to be a wayward adolescent, full of conflicts, with an unintegrated personality—now sanguine, now aggressive, now confused. In so far as he has a unity, this derives from his preoccupation with national and 'school' differences in outlook, on the question of personality theory.

The book consists of 22 chapters each by a different writer. Eleven are listed in the Introductory Note as psychologists from the United States, the other half being drawn from Belgium, Canada, England, France, Germany, Holland, Italy and Switzerland. Despite the evident efforts of the contributors and the editors towards topicality, the reader increasingly gains the impression that *plus ça change plus c'est la même chose*. The changing and occasional coining of 'technical' terms merely emphasize the inevitable recurrence of the concepts. There is much play with the traditional antithesis between 'philosophical' (or speculative, deep, characterological, understanding—or European, i.e. German) on the one hand and 'scientific' (or experimental, objective, behaviouristic, explanatory—or American) on the other. One of the useful functions of the book, however, is to demonstrate implicitly how untenable, as well as sterile, such dichotomies are. First, it is obvious from contributions on both sides of this straw fence that clarity and experimentation are not the prerogative of the 'scientific' psychologist, nor obscurity and special pleading the hall-mark exclusively of the philosophically-minded. Secondly, a glance at the names and present home of the contributors points the absurdity of any classification nowadays based on geographical or national considerations. The United States are represented, for instance, by Else Frenkel-Brunswik and Werner Wolff just as they are by Gordon Allport and Frederick Wyatt; the two U.K. contributors are Hans Eysenck and Cyril Franks; Otto Roser hails from Italy, Wolfgang Luthe from Canada.

The book is divided into six sections entitled respectively: Overview, European Trends, Theory, Methodology, Commentary and Resources. The Overview consists of Allport's chapter. This is mainly a discussion of what he calls the 'Anglo-American' and 'Continental' approaches. 'Resources' comprises a useful, well-classified Bibliography (324 references) by David and a Note on the International Congress by MacLeod.

Of the four central sections which form the body of the book, the reviewer found most stimulating the chapters by van Lennep, 'Projection and Personality', Wolff, 'New Concepts in Experimental Depth Psychology', and Frenkel-Brunswik, 'Perceptives in Psychoanalytic Theory'. These three illustrate the practicability of combining 'understanding' with 'explaining' and even 'predicting'. There are also two interesting chapters in the Commentary by, respectively, Wyatt and McClelland. In spite of the bird's-eye view which these two writers attempt, they manage to be both constructive and tolerant.

A. W. HEIM

Psychology in the Soviet Union. Edited with an Introduction by BRIAN SIMON. London: Routledge and Kegan Paul. 1957. Pp. viii + 305. 32*s*.

Two years ago, after a visit to the Soviet Union which included discussions with several Soviet psychologists, Mr Simon decided that there was need for a book in English containing samples of recent Soviet work. He put this suggestion to Prof. Smirnov, the Director, and Prof. Menchinskaya, the Deputy Director, of the Institute of Psychology in Moscow, and they immediately responded by sending him thirty papers recently published by Soviet psychologists. The present volume contains edited translations of twenty-two of these papers, together with a concise and illuminating survey of the whole field by Mr Simon. None of the papers is more than 6 years old, and all but four were first published during the last 3 years.

The papers are arranged in five main groups, according as they have been judged to be primarily concerned with: the general direction of research in Soviet psychology; Pavlovian findings and principles; perception and memory; learning and development; or general psychological theory.

While the papers are valuable for their detailed discussions of particular subjects, they also provide a general picture of current Soviet psychology. Some of its main features may be briefly indicated. First, orthodox Soviet work is firmly based on the principle that mental processes and behaviour must be explained by reference to the neural processes that underlie them. This is 'the only correct view' (p. 29), and although some recent articles have deviated from it, they 'aroused keen opposition from the general body of Soviet psychologists' (*ibid*). Secondly, the development of an individual is assumed to be determined by his 'conditions of life' and his 'activity'. The influence on development of the 'conditions of life' must not be understood in 'an over-simplified mechanical way' (this is 'the error of the behaviourists'): it involves 'the whole complex dynamics of the cortical processes' (p. 47). Thirdly, all consciousness is held to be a reflexion of the objective world, including, in Pavlov's words, 'the entire organism and all its component parts', and it arises as the result of processes of analysis and synthesis at various levels in the nervous system, and especially in the cerebral cortex. (But, strangely enough, there is no reference in these papers to information theory—or, for that matter, to cybernetics.) Fourthly, Pavlov's theories, and even his *obiter dicta*, are usually accepted as authoritative, and, in the explanation of various psychological processes, there is considerable uncritical use not only of conditioned reflexes but also of the 'second signal system' and other mechanisms involved in his account of 'higher nervous activity'. Fifthly, mental tests and all other psychometric techniques for assessing individuals are roundly condemned. It is charged against them that they presuppose fixed differences between individuals, that they are unreliable, and that they are 'restricted to a mere statement of the result, without disclosing why or how it was achieved, or what prevented it from being better' (p. 189). Sixthly, except in their reports of international congresses, Soviet psychologists rarely refer to 'the psychology of the West', and, when they do refer to it, it is usually to condemn psychoanalysis and Gestalt psychology for their subjectivism, behaviourism for its over-simplified mechanist standpoint, neobehaviourism for its neglect of psychoneurology (and its undifferentiated use of 'habit'), and 'so-called social psychology' for its 'superficial inquiries', which have 'no scientific significance whatever' (p. 295). Finally, Soviet psychologists, like Russians in other fields, practise severe self-criticism—self-criticism, that is, in terms of the orthodoxy to which they are expected to conform. Thus Prof. Smirnov summarizes (pp. 41–5) nine 'important gaps and deficiencies' in Soviet research. He says, among other things, that 'theoretical work is inadequate', that 'there is an almost complete lack of studies of the psychology of labour', that 'there has been a significant reduction in work on psychopathology', and (as we have heard of psychology nearer home) that 'the scientific level of research is not always satisfactory: some investigations have no relation to either theoretical or practical tasks; others are superfluous; certain conclusions sound like well-known truisms'.

Mr Simon tells us that in the Soviet Union psychology is taught to all pupils during the final year of their 10-year schooling, that it is available to undergraduates in 'the relevant departments' of the universities, that it forms part of the course for all intending teachers in universities and training colleges, and that it is developed both in special research institutes of psychology in Moscow, Leningrad, Kiev and Georgia (presumably in Tbilisi) and in research institutes in education in all the republics. The papers in this book all come from Moscow, Leningrad or Kiev, but they sometimes refer to psychologists in other parts of the Soviet Union (though not to the Georgian followers of Uznadze, who are said by *Soviet Culture* to be involved in heterodox excursions into 'Freudism'), and we may infer that, although most Soviet psychology is to be found in these three cities, there is also something, but not perhaps very much, going on elsewhere.

In all the papers there is evidence of the influence of the ideological strait-jacket. Everything is made to conform strictly to officially approved Pavlovian principles; there is explicit submission to 'the Decree [regarding the influence of heredity and environment on the development of an individual] of the Central Committee of the Communist Party of July 4, 1936' (p. 47); there is repeated reference to the sacred texts—the 'classics of Marxism-Leninism' (p. 133)—and to such particular Marxist doctrines as that 'sensations, which are inseparable from man, developed historically in the course of development of social labour' (*ibid.*); and, in a paper first published in 1954, there is solemn acceptance of Stalin's pronouncements on the nature of language and its relation to thought. Again, while valuable advances are being made in psychoneurology, which Soviet psychologists sometimes call 'the neurodynamic explanation of mental processes', much of the research in educational and child psychology (and there is apparently more research in these fields than in any other) is concerned with specific, practical problems rather than general, basic issues. Still, even here, Soviet psychology has important things to say about the role of learning, and of social factors, in perception, motivation and the development of personality; and, as these matters have only recently become prominent in British and American psychology, it is useful to see what is being made of them in the Soviet Union, where the ideological climate has long led to their being among the principal subjects of study.

Clearly, this is an informative and valuable book, even though the value of the collected papers perhaps lies less in their contribution to psychological knowledge than in the picture they present of the main characteristics and tendencies of Soviet psychology as it was 2 or 3 years ago.

REX KNIGHT

Psychological Tests and Personnel Decisions. By L. J. Cronbach and G. C. Gleser. Urbana: University of Illinois Press. 1957. Pp. xii + 165. $3.50.

Using the concepts and techniques of Decision Theory the authors tackle the problem of selection and placement with the most interesting results. The term selection covers not only ability tests but also interviews and personality tests; placement refers to educational, industrial, and psychiatric placement. The analytic device used in this application of Decision Theory is the comparison of strategies, i.e. the policies used to select and place people. The object is to choose the strategy that is most beneficial to the institution concerned. By using pay-off functions the authors show how it is possible to compare the utility of training programmes over the whole range of a given ability when the conventional methods only give the best programme without any information on the usefulness of the programmes for different levels of ability. For developing training schemes such detailed knowledge would be invaluable in making policy decisions on the relative advantages of grouping people in terms of ability. Perhaps the most important contribution this new approach makes is to query the adequacy of the present methods of validating tests. The coefficient of determination is found to apply only under certain conditions and the coefficient of forecasting efficiency is suggested as having no relevance whatsoever. The present methods of validating tests do not take into account what the authors call the *a priori* strategy, i.e. knowledge such as academic records which could be used as a means for selection and placement without reverting to tests. They maintain that it is in relation to this that a test should be validated. Furthermore, before deciding to use a test the advantages which would accrue from adjustments to training programmes and job simplification must be considered since they are shown to be intimately associated with the value of the validity coefficient. How much a test should be lengthened to increase its validity must be balanced with the increased cost of the test; equations are given for finding the optimum length of a test for this condition. The authors conclude there is no single validity coefficient; a test must be considered in relation to the objectives it has to attain. Selection using single and multiple tests is analysed to provide equations for determining the optimum cut-off scores and selection ratios. Another important question discussed is whether to measure a small number of aptitudes efficiently or a large number with moderate efficiency. The problem as seen by the authors is a matter of balancing the two extremes to obtain maximum utility from the tests. Again the problem of validity is encountered and it is pointed out that the contribution to utility of a number of moderately efficient tests may be greater than a smaller number of highly valid tests. Classification problems are taken as the general case of selection and placement but the topic is not developed because it entails numerous restrictions. To evaluate the outcome of a strategy requires partly or wholly subjective judgments, and is, therefore, open to criticism. This weakness is no greater than that of other psychological theories; on the contrary the explicit statement of outcomes makes it stronger than most.

This stimulating book has brought much needed cohesion to the subject and has overcome some of the problems that have faced conventional test theory. Whether the solutions it advocates will be accepted by its readers will depend not upon the rigour of the mathematics, which cannot be faulted, but upon the assumptions from which the mathematical arguments are developed. The generality of the new approach will appeal to the mathematically inclined psychologist but the consequences of it are important to all applied psychologists. The empirical work entailed in the development of tests in accordance with the theory will be more exacting but at the same time more rewarding than that for conventional theory. Decision Theory is potentially a powerful tool for exploring new areas of psychology. To apply it, however, requires mathematical competence and a sound knowledge of the subject-matter to be explored; both qualities are shown by the authors.

For applied psychologists this book is an absolute must because it deals with classification problems pertinent to their work. It can be recommended to all psychologists as being amongst the studies extending psychological theory into new and important fields. Whilst at times non-mathematical readers will find the arguments a little difficult to follow they need not be abashed by the book since the authors put the minimum of formulae in the text, preferring to give the mathematical development of the theory in appendices. Before reading this book most readers will find it helpful to read *Design for Decision* by I. D. J. Bross (New York: Macmillan Co., 1953) which gives a good popular introduction to Decision Theory. There are two misprints in Appendix 3, p. 142; they concern formulae (3·3) and (3·4).

H. C. BAKER

Information Theory: *Third London Symposium*. Ed. by COLIN CHERRY. London: Butterworth. 1956. Pp. xii+401. 70*s*.

A book of 400 pages presenting papers on forty different topics confronts the reviewer with something of a problem. Information Theory is not a tidy and coherent body of systematically related concepts, nor is it a great inductive generalization from convergent experimental studies. Or at any rate if it is either of these things the fact does not emerge from this symposium. A hostile reviewer would accordingly have an easy task, by selective quotation, in making mincemeat of the whole project. But he might thereby reveal more of his own limitations than of the significance of the work in question. It is probably not unfair to say that Information Theory represents a movement of youngish men, even though the pioneers are now venerable. There is vigour, there is enthusiasm, there is scant regard for the established inter-subject boundaries of the past, there is preference for the fertile suggestiveness rather than formal conclusiveness, there is reliance on teamwork and mutual intellectual support—the characteristics, in fact, of Operational Research. The social psychologist and the anthropologist will discern a number of familiar tendencies in such a movement—it has features in common with a nascent religion.

The hardest aspect of the movement for those reared in a slightly earlier discipline is not its mathematical dressing nor even its conceptual complexity but the ambiguity which saturates every discussion. If you have regularly had your knuckles rapped for loose language, whether in logic or in linguistics or in methodology or in philosophy, it is hard to adapt yourself to an atmosphere in which quite other standards prevail. It is like being brought up in a society adhering to strict matrimonial orthodoxy and moving into one where free love is accepted. You may try to adapt, you may go through the motions of a new conformity, but your conscience makes you miserable. And if you have a sneaking belief that the new society is, in important respects, better than the old your distress is all the more acute. These information-theorists may be swimming in a sea of ambiguities but they seem to be getting somewhere.

It is important not only to assess what Information Theory is saying but, even more important perhaps, to appreciate the innovation in methodology implicit in the fact that so many very different things are being said in what purports to be a single context. It is a context in which speech jostles with thermodynamics, chance with electronics, librarianship with arithmetic, games with linguistics, taxonomy with optics, perception with induction, logic with neurology—in fact almost anything goes. And out of this higgledy-piggledy empire we are getting, among other things, automation. Its own conceptual unity, and, in time one hopes, its own coherent discipline must take shape out of its seemingly amorphous assemblage of diverse elements.

Information Theory is already making piercing inroads into most of the existing major disciplines—into logic and mathematics, physics and biology, linguistics and psychology, medicine and technology. Let us not underestimate the magnitude of its impact. It is unlikely that anyone can as yet correctly interpret its significance. It is akin to one of these shifts of ecological equilibrium in which the interacting factors are so numerous, and so resistant to precise calculation, that the outcome is quite unforeseeable. We will each look for what interests us most in the phenomenon. The practical effects are more immediate and more obvious than the long-term theoretical implications. One of these latter is surely the methodological discovery that ambiguity is not only inevitable in all language but that it may be a principal source of novel and usable concepts—just as free love might break up the genetic stratification of a rigid social system in which caste separates marriage-groups as effectively as geographical distance. Novel genetic clusters will emerge, and similarly Information Theory is throwing up quite novel conceptual clusters.

The importance of ambiguity lies in the light that it throws on interpretation. We have been obsessed too long with an ideal of one-one correspondence between elements in the stimulus and elements in the percept. The realization that stimulus-elements serve as cues for the evocation of interpretative framework is not novel but its implications go much further than the theory of perception has indicated. The psychology of productive thinking is involved here, and with it the psychology of communication. The terminology of 'signal' and 'noise' borrowed from telephone engineering is useful here and has been put to good effect in this symposium by D. J. Bruce of Reading in a simple experiment capable of considerable development. Four series of tape-recorded words are presented a number of times to a group of subjects. They are combined with noise at varying levels, greatest at the first presentation and progressively diminished, with of course a resulting increase in word-recognition. Two of the word-series were of a random character. Each of the other two was selected from a specific context namely, *parts of the body* and *kinds of food*. The outcome, in terms of graphs of recognition-performance, is, of course, what one would expect. The interest lies in the detailed light thrown on the recognition-process. We tend to speak of 'context' as if it were something appertaining to words. When we appreciate the fact that the context must inevitably be supplied by the subject we see the storage-system of the brain in a new light. It not only stores 'bits of information' but a hierarchy of established contexts. And it must possess mechanisms for switching freely into one context after another in its groping after interpretations. It is perhaps unfair to select for mention a single paper out of a symposium of forty papers but there is not only the excuse of psychological interest and of limitations of space but the further justification that this experiment, rightly interpreted, will be found to exemplify most of the major concepts occurring in the Theory of Information.

G. PATRICK MEREDITH

On Human Communication. By Colin Cherry. New York: Wiley. London: Chapman and Hall. 1957. Pp. xiv + 333. 54*s*.

This book is intended as a general introduction to a series of forthcoming texts by various authors on specific aspects of communication. It therefore touches on each of the sciences connected with the subject, in a series of self-contained chapters: one introductory, one historical, and the remainder divided between linguistics, phonetics, statistical communication theory, semantics, and psychology. Each chapter is meant for readers who are not specialists in the particular topic, so that the book may well attract those who want an outline of the ideas of neighbouring disciplines. A number of references are given, which allow further exploration: their value is reduced by the absence of an author index or any other way of finding the portion of text in which a particular reference is cited.

It would be easy for any professional psychologist to find fault with the chapter on that subject: certainly those of, say, the learning theory variety should be warned against regarding the last chapter as a comprehensive guide to the activities of their cybernetic colleagues. But such criticism misses the point of the book. We ought rather to ask whether it succeeds in communicating ideas across the barriers between the sciences. The strongest parts, as might be expected, are those dealing with communication theory and signal analysis. On these topics more is expected of the reader than in other chapters: he can hardly survive if he is frightened of an integral sign, and many psychologists are. But easier short treatments are already available; the merit of this one is rather that it is intermediate, going into some detail without the length and complexity appropriate to an engineering handbook. The other disciplines are handled adequately on a more elementary level.

All in all, this is probably the best introduction to all these neighbouring fields yet available in a single volume. There are few people who could have written it single-handed. However, Dr Cherry's broad sympathies are only too clearly revealed by his remark that 'the mechanical engineer designs a motor-car to operate with its driver, as an integrated man-machine unit'. Too often the engineer does nothing of the sort, and the car-driver suffers. Equally many psychologists are content to regard language as merely a kind of bee-dance. In this case it is psychological theory which suffers. Books of this sort serve to remind us that there is more behaviour in our daily life than is dreamed of in our philosophy

D. E. BROADBENT

Interviewing for the Selection of Staff. By E. Anstey and E. O. Mercer. London: Allen and Unwin. 1956. Pp. xiv + 111. 10*s*. 6*d*.

Two experts have collaborated to produce one of the few really effective aids to those who, often without qualifications and training, have to do a job in the field of applied psychology. It is a popular account in that the layman will find it easy to read but it is certainly not popular in the pejorative sense of that word. There is nothing in it that a university psychologist would not be proud to have written.

After modestly limiting the purpose of the book to those who are occasionally called on to interview people for appointments or promotions, a few pages are devoted to the aim of the selection interview. The next chapter advises the interviewer on how to ask himself questions before interviewing, questions about aspects of the job and then questions about the candidates to which he might hope to discover answers by means of the interview. A short chapter on 'Preliminaries and Setting' bears the stamp of the authors' experience. A miracle of condensation is achieved in the heart of the book in a 10-page chapter called 'Method during Interview'. It could not be bettered in such a small space.

Excellent advice is offered on note taking and report writing, and specimens of rating scales and interview forms are shown. Critical and knowledgeable accounts are given of Board Interviews and Multiple Interviews and good sense is talked about hauling all the material together into conclusions and recommendations. A novel feature is a chapter entitled 'Advice to a person about to be interviewed'—and good advice it all is. Here is a sample. 'You may be asked about your family. Whatever your origin, refer to it openly and without defensiveness. If your father had a distinguished career, why conceal the fact? If he was poor or for a time unemployed, you may be the more able to pay tribute to the way your parents brought you up.' For good measure three appendices and a booklist are added. A bargain at 10*s*. 6*d*.

DENIS MCMAHON

Supervisory and Executive Development: A Manual for Role Playing. By N. R. F. MAIER, A. R. SOLEM and A. A. MAIER. New York: Wiley. London: Chapman and Hall. 1957. Pp. xi+330. 52*s*.

For those who wish to indulge in role playing as part of human relations training, this book provides complete instructions. Its authors claim that it is equally applicable to industrial executives and supervisors, individually or in small or large groups, and to the family circle. No skilled instructor is required, for it does not set out to be a text-book; it 'proposes to avoid giving answers. Some principles and facts may be introduced, but basically the aid that may be supplied is in the form of raising questions and exploring issues'. The course of twenty case-studies covers the ground from straightforward, everyday problems met by first-line supervisors to complex organizational questions confronting managers and directors. Eight of these studies appear to consist of new material, the remaining twelve being adaptations from work published previously by the senior author. Only one case requires more than eight players, and six of them contain only two parts. However, minutely detailed directions show how groups of the order of a dozen up to several hundred members may derive simultaneous benefit from the programme. For each case there is an introductory statement focusing the problem to be solved. Next come instructions for the conduct of the 'play' (either Multiple, or Single Group, Role Playing Procedure, including sometimes the Skit Completion Method or the Dramatized Case Method), followed by background to the specific study and roles for the players. Finally, there is a discussion about the general human relations principles involved.

It will be clear from this account that the book is likely to be of value to serious students of psychology and sociology only in so far as they may wish to familiarize themselves with these particular applications of role-playing techniques; it is not written for them. Its potential field of usefulness includes, rather, the various extra-mural, W.E.A., technical college and adult education centre courses concerned with human relations, and training departments in industrial organizations. Except possibly for the last of these, the price will be a serious drawback, since each member of the group requires a copy; the roles are deliberately not available separately, and reproduction is expressly forbidden. For groups of all these kinds, some rewriting would be needed to make the settings more realistic, and the names and vocabulary more familiar (the word *unique*, for example, occurs at least four times in the first 10 pages). In fact, the book may well be of more value as a starting point for an instructor than as a manual for class members.

On the one hand we may feel grateful to the authors for having given so careful a picture of one aspect of contemporary American supervisory training practices; on the other, we are left with regrets that no evidence is included as to the effectiveness of role-playing techniques in furthering one of the most important aims of such training—the change of attitudes in actual work situations. Nor is the need for such evidence mentioned.

DERMOT STRAKER

Psychology of Personnel in Business and Industry. By ROGER M. BELLOWS. London: Staples. 1956. Pp. xii+467. 42*s*.

This thick book is the first British publication of the second American edition. It is written more for business executives and personnel managers than for psychologists but although briskly written is not at all a popular handbook. Part one is a very interesting but, to British readers, not very useful layout and description of the kinds of work different kinds of personnel people do in U.S.A. Part Two, about one-third of the book, reports research on workers' satisfactions, group dynamics and attitude surveys and covers in a condensed but effective way the material used by Viteles in his *Motivation and Morale in Industry*. This Part has grown and has been brought up to date since the first edition. Those whose acquaintance with the literature on the theory of social and industrial psychology stops at Roethlisberger, Lewin and Moreno will find this part of the book very useful.

Part Three is an account of modern selection methods and of some aspects of training. In this Part the most interesting chapters for this reviewer were one on Recruiting, which sets the whole problem of selection in a wide context, and another on Readability of training material in which much more than Flesch's work is reported. (I particularly liked the formula for calculating the 'Fog Index' of a piece of writing.) After finishing the book one has a fairly clear picture of the author: he seems to be an efficient, keen, executive-type industrial psychologist who is research minded and a scientist at heart. The book pretends not to be scholarly (presumably so as not to put off the kind of reader for whom it is written) but it is packed with research material and does not stray beyond the evidence; one often discovers that the author has himself carried out applied research in a particular field although some of it has not been published. Psychologists working in the social-industrial fields will find it a useful book, students interested in psychology applied to industry should have access to it and all British personnel managers would find it very valuable.

DENIS MCMAHON

Intelligence in the United States. By JOHN B. MINER. New York: Springer. 1957. Pp. xii + 180. $4.25.

This book is a strange mixture of over- and under-statements. The author claims that he has made a survey of a cross-section of the U.S. population 'with conclusions for manpower utilization in education and employment'. This would appear to be overstating the case since the book is based simply on the results of testing 1500 subjects (1347 whites and 153 negroes, ranging in age from 10 years to 75+) on one 20-word, multiple-choice vocabulary test, and comparing their scores with their occupations ($n=316$), race ($n=2$), religion ($n=4$), social class as self-estimated ($n=4$) and sex ($n=2$). The other variables considered include education, place of residence, marital status, and age. On the basis of his subjects' scores on the vocabulary test and their position on these other variables, Dr Miner estimates what proportions of people in the U.S.A. are over- and under-employed, over- and under-educated. He expounds, in percentages correct to one—sometimes two—decimal place, how many should be 'promoted' and how many 'demoted'. It is among these conclusions especially that understating of a kind occurs. For example, having ascertained 'the degree of manpower wastage as indicated by tests of verbal learning' and specified the numbers at every level who should be promoted, he adds with apparent light-heartedness: 'In order to determine the extent to which our present occupational placement system produces a deviation from one *based entirely on a verbal ability criterion,* all that is necessary is to compare an individual's vocabulary score with the ranges that have been established and then determine whether it would be necessary to promote or demote him to ensure his employment at a level consistent with his *intelligence.* The degree of deviation can then be specified in terms of the percentage of the working population who would have to be promoted to a higher occupational level to attain a perfect intelligence-occupation correlation. An equal number will, of course, have to be demoted in order to keep the number of people working at each level constant' (reviewer's italics). The last sentence (which ends the section) strikes the reviewer as a masterly understatement of difficulties—leaving aside any question as to the defensibility and value of Dr Miner's plan.

Dr Miner is interested above all in avoiding wastage. Towards the beginning of his chapter on 'The utilization of intellectual resources in the educational system' he says: 'The data, to be presented in this and the following chapter, on educational and occupational manpower wastage are unique.' The reader, pursuing page after page of what can only be called autistic thinking, becomes inclined to agree. Dr Miner's conclusions rest on two assumptions. The first is the infallibility of the 20-word vocabulary test at all levels and for all purposes. The second is the fact—rather than assumption—that 90% of any given group score above the 10th percentile! Lest it be felt that this analysis is less than fair to Dr Miner, the following two excepts are offered for consideration. (*a*) 'A score of seven comes directly above the tenth percentile point of the semi-skilled workers, and accordingly anyone with a score of seven may be considered as capable of semi-skilled work but not of skilled or highly skilled jobs. Anyone with six or less words correct is, by exclusion, intelligent enough only for unskilled labor.' (*b*) 'In establishing minimum requirements for performance at each educational level, the score directly above the 10th percentile was chosen.... In using the 10th percentile point as a lower bound, a tendency to underestimate the number of students who could do more advanced work is introduced.'

The remaining parts of the book deal primarily with the findings of other workers who have investigated test scores as related to school, home and occupational environment; and with 'social stratification', the main variables being race (black and white) and social class. On the last two points, Dr Miner makes several naïve comments. For example, referring to social class he says: 'What leads people who do not meet the objective criteria for membership to identify with the upper class is impossible to say.' Towards the beginning of the book the reader has difficulty in deciding just where the author is going. By the end of the book, he has decided that it would have been better not to travel at all than to travel hopefully.

A. W. HEIM

Perceptual Processes and Mental Illness. By H. J. EYSENCK, G. W. GRANGER and J. C. BRENGELMANN. Maudsley Monographs No. 2. London: Chapman and Hall. 1957. Pp. 144. 25*s*.

The inquiry described in this monograph was undertaken in order to investigate the possibility that the perceptual processes of neurotic and psychotic individuals differ in some way from those of normals; to 'isolate differentiating variables and offer explanatory hypotheses concerning the processes responsible for differentiation'. Such an investigation could be of great interest and value, particularly if the nature of the differences could be determined. Unfortunately the experiments carried out by the authors of the monograph cannot advance our knowledge greatly. They presented a large and heterogeneous collection of tasks to groups of normal, neurotic and psychotic individuals. These included sensory tasks (dark adaptation, colour vision, acuity, after-images, etc.); oculo-motor tasks (binocular fusion, accommodation, etc.); perceptual tasks (Gottschaldt figures, visual illusions; number, pattern and object perception, etc.). In so far as can be judged from the brief and not very lucid descriptions given, the experiments were frequently carried out by techniques too crude to show up other than gross differences of function; they were not always adequately controlled; the readings were too few; and the scores were in many cases expressed in arbitrary units the physical values of which are not given. All we can conclude is that in many of these tasks the neurotics and psychotics performed slightly worse than did the normals, but that there was a great deal of overlap between the different groups for most of the tasks. An attempt was also made to show from some of the experimental results that a classification of the subjects in terms of Jaensch's 'integrate-disintegrate' typology gave some indication of a stronger tendency towards the 'disintegrate' type among abnormals than among normals. But the results were not very consistent, as is usually the case with perceptual typological classifications of this kind.

The scores of the normal group of subjects in the perceptual tasks and on certain non-perceptual tests (intelligence, motor, physiological) and questionnaires were correlated and submitted to factorial analysis. Five factors were extracted, the third of which had high loadings on visual acuity (as measured by the Snellen chart), fluctuation of after-images, perception of letters and of objects. In these the distribution of scores of normals and abnormals shows a comparatively small overlap, so that about three-quarters of the cases in the two groups were correctly differentiated. However, it is doubtful whether even these tasks could be used as diagnostic instruments as they stand.

A far higher degree of care and accuracy in the design and performance of the tasks, such as is normally used in carrying out experiments of this kind, is obviously necessary before any valid data can be obtained to demonstrate that neurotics and psychotics perceive differently in any consistent way from normals. Inter-correlation of the scores of the former in a number of different perceptual tasks might give some indication of this. But also the type of experimental task and the results obtained must be *psychologically* investigated and analysed before anything can be deduced as to the nature of the differences. Finally, it is probably important to take into account the effects of incentive conditions, since it is quite possible that abnormal subjects are poorly motivated to undertake tasks which are tedious to perform accurately and carefully, rather than inferior in perceptual capacities.

M. D. VERNON

Left handedness: Laterality Characteristics and their Educational Implications. By MARGARET M. CLARK. University of London Press. 1957. Pp. xvii + 214. 15*s*.

This monograph, sponsored by the Scottish Council for Research in Education, is divided into two parts, the first part reviewing the literature and the second reporting the results of an investigation into the hand, foot, eye and ear preferences of children drawn from eight Glasgow primary schools.

The literature on handedness and allied topics is extremely diverse, and the author has done well to provide such a concise and judicious summary of previous work. The explanation of hand dominance falls under two headings: granted a division of labour between the two hands, why has man become predominantly right-handed?; and granted this right-handedness, why does a select minority insist on deviating towards the opposite extreme? These two riddles remain unsolved, yet some ingenious theories have been put forward, most of which receive honourable mention here. A chapter is devoted to the genetic approach, and the development of laterality in young children is briefly outlined. The possible connexion between enforced change of handedness and stuttering, and between crossed laterality and reading disability is critically examined, while the special difficulties that are apt to beset the left-handed writer are given due prominence.

The author's sample of 330 children aged between 11–12 was unselected as regards laterality, and free from any known bias in respect of intelligence and social background apart from being entirely urban. A battery of 18 tests, most of them designed to measure preference rather than skill, was applied to each child individually, and a writing test was administered in the classroom. The results are analysed with painstaking attention to detail; and while it may not be easy for the general reader to keep his perspective as he gropes through a forest of statistics, the raw material is all there for anyone who wishes to discover how many left-footed hoppers were observed to kick with the right foot, or the ratio of left-eyed to right-eyed peepers.

Some of the conclusions drawn, e.g. that left-handers can hold their own both as writers and as scholars, will come as a relief to some parents, although probably as no surprise to workers in the educational field. When the author concludes finally that 'no essential difference was apparent between left-handers and right-handers, except for their use of a different hand' one wonders whether the matter ought really to have been allowed to rest there, for surely their behaviour in the test situation cannot be expected to tell the whole story. Indeed at least one reader, himself a member of that 'sinister minority' to which the author is evidently proud to belong, is doubtful whether any valid conclusions can be drawn about the significance of left-handedness from a sample which included no more than 18 left-handed writers. Yet while one might have wished for a more imaginative approach to the problem, a survey of this kind has nearly always something to contribute. All who are professionally concerned with primary education would do well to read it.

MICHAEL HUMPHREY

Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf. Ed. with an Introduction by John B. Carroll. London: Chapman. New York: Wiley. 1956. Pp. xi+278. 56*s*.

Benjamin Lee Whorf was an industrial chemist, trained at the Massachusetts Institute of Technology, who attained a remarkable reputation as a 'trouble shooter' for a Fire Insurance Company. Early in his comparatively short life—he died at forty-four—he turned his unusual abilities to the study of linguistics, working for a while with Sapir at Yale. Apart from purely linguistic essays this collection of his writings contains a number of papers maintaining the thesis that 'thinking', in particular thinking about such ultimate metaphysical problems as the nature of space and time, derives from the language used by the thinker. The general thesis is not of course by any means new, but its derivation from American Indian languages such as Hopi will be interesting to psychologists. One reads the book with respect for the power and originality of the author's mind, but with the feeling that the psychological issues are treated from the outside, by one only superficially cognizant of psychological method and data.

GEORGE HUMPHREY

A New Approach to Figure Drawing. By L. Caligor. Springfield, Illinois: Thomas. Oxford: Blackwell. 1957. Pp. xii+148. 34*s*. (Test materials available at 45*s*.)

Projective techniques utilizing human figure drawing are becoming more widely known and the present book adds another method to the approaches associated with the names of Goodenough, Wechsler and John Buck. The new technique is intriguing—the subject goes from one drawing to the next whilst the preceding drawing is visible to him through the transparent paper; he may trace it and virtually repeat it, modify it or ignore it completely His handling of this situation, as well as the various details of the drawing, the placement and line quality are considered valid diagnostic clues.

Dr Caligor provides a scoring system based on objectively scorable dimensions and attains a high interscores agreement. The scores are stated 'in terms of a deviation from a statistical norm approximately the mode scores of normal subjects'. Unfortunately the book contains no statistical material and the only relevant reference given by the author refers to a study of items differentiating between 21 college males and 21 hospitalized paranoid schizophrenics—this could scarcely be called a normative study serving as a basis for a complete test manual to be used in clinical practice.

Even more disappointing is the systematic list of suggested interpretations for the various scoring categories which are, no doubt, based on wide clinical experience, but which would be more convincing and valid if supported by some evidence. Though Dr Caligor's technique shares this fault with the systems devised by his predecessors, one feels that nowadays, having demonstrated the wealth of clinical material obtainable from drawings, investigators should approach this field more systematically and not by-pass such important factors as sex, age, intelligence, cultural background, etc.

In the meantime, this new publication's main merit is the presentation of a novel and promising projective technique and of a scoring method which may well be used for testing experimentally the validity of drawing interpretations.

H. C. GUNZBURG

On the Early Development of Mind. By EDWARD GLOVER. London: Imago. 1956. Pp. xi + 483. 45*s.*

Dr Glover is a veritable Mrs Battle. He is all for a clear fire, a clean hearth, and the rigour of the Freudian game. He regards most of the neo-Freudian developments much as that good lady would have contemplated Bridge or Canasta, or any other sectarian variant of her favourite sport. 'The object of the following survey'—his 1949 paper on *The Position of Psycho-analysis in Great Britain*—'is to promote a clear distinction between the standard findings of psycho-analysis and those more speculative accretions which...do more credit to the hearts than to the heads of their sponsors.'

This first volume of Selected Papers on Psychoanalysis is, in virtue of the omissions, as prodigious as the visible peak of an iceberg, and much of it does not make easy reading. It includes 28 papers out of the 177 that Dr Glover has published over a period of forty years, and that are listed in full in a personal bibliography at the end of the book. This selection stretches chronologically from 1942 (*The Significance of the Mouth in Psycho-analysis*) to 1954 (*The Frontiers of Psycho-analysis*), and is amplified, garnished, and enlivened by introductions and notes—expository, hortatory, critical, topical, and controversial—on Eysenck, Goldman-Eisler, the relations between the different disciplines bearing on psychology, the risk of the dilution of psychoanalytic values in the course of inter-disciplinary team-research (not that Glover is opposed to this), and a variety of other matters. Inevitably, the papers are not all of equal significance to the general psychologist, and not all of them, despite the title of the book, and its reiterated emphasis on the value of the developmental approach in clinical practice and theoretical discussion, are equally important to or likely to be understood by the genetic psychologist. But orthodox psychoanalysis, in one of its most significant aspects, is an attempt at a genetic psychological theory, and it is important that the child psychologist should be clearer than he often is about the origin and nature of the formulae that he employs without adequate examination. Yet here, in reading these papers, we are reminded that, whatever his claims, the analyst is in no better case than the experimentalist, like Michotte or Hebb, whose formulations cry out for genetic underpinning. 'The obscure stages of early mental development which are not accessible to direct analysis...can be outlined only by means of plausible "reconstruction"', which exposes the analyst to the dangers of 'encroachment by speculations which, uncontrolled, are little more than wild surmises'. These are Dr Glover's words, and it is this continuing element of 'wild surmise' that still supplies much of the detail of the psychoanalytic account of early psychological development.

But there must be few psychologists, whether they are working in the clinical field or not, who would fail to find stimulation in this collection of essays by a clinician of experience and wisdom. These qualities lend force to the slighter and less technical papers, some of which are among the most interesting in the book; and it is these too, rather than the weight of psychoanalytic theory, that give point to many of Dr Glover's *obiter dicta* as, for example (from *The Unconscious Functions of Education*): 'even amongst analysts the view that education is an *unconscious* mechanism...serving amongst other purposes the purpose of *unconscious instinct inhibition* is rarely emphasised', or again from a critical notice (*Psychology and the Public*) written in 1945: 'since defective children are almost invariably handled with unwisdom, they naturally respond by retreating into their lonely interiors, from which it takes a good deal of love and reassurance to entice and rescue them.' The present writer holds the view that there is urgent need for a common understanding, not merely between psychiatrists and professional psychologists 'on the job' but at the level of principle between education, psychology, and medicine (paediatrics, psychiatry, and psychoanalysis), and there is much support for that broad proposition in this record of the development of Dr Glover's thinking over many years of practice, reflexion, and astonishing literary productivity.

GEORGE SETH

CORRIGENDUM. In the review of *Interpersonal Diagnosis of Personality*, the last sentence of the penultimate paragraph on p. 231 should read: 'The Kaiser Foundation Research may well have a like influence on clinical and social studies during the next decade.'

BUREAU OF EDNL. & PSYL.